

H. Russell Bernard

**FOURTH
EDITION**

**RESEARCH
METHODS IN
ANTHROPOLOGY**

**QUALITATIVE AND
QUANTITATIVE
APPROACHES**

**RESEARCH
METHODS IN
ANTHROPOLOGY**

RESEARCH METHODS IN ANTHROPOLOGY

FOURTH EDITION

*Qualitative and
Quantitative Approaches*

H. Russell Bernard



ALTAMIRA
PRESS

A Division of

ROWMAN & LITTLEFIELD PUBLISHERS, INC.

Lanham • New York • Toronto • Oxford

ALTAMIRA PRESS

A division of Rowman & Littlefield Publishers, Inc.

A wholly owned subsidiary of The Rowman & Littlefield Publishing Group, Inc.

4501 Forbes Boulevard, Suite 200

Lanham, MD 20706

www.altamirapress.com

PO Box 317, Oxford, OX2 9RU, UK

Copyright © 2006 by AltaMira Press

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior permission of the publisher.

British Library Cataloguing in Publication Information Available

Library of Congress Cataloging-in-Publication Data

Bernard, H. Russell (Harvey Russell), 1940–

Research methods in anthropology : qualitative and quantitative approaches / H. Russell Bernard.—4th ed.

p. cm.

Includes bibliographical references and index.

ISBN 0-7591-0868-4 (cloth : alk. paper)—

ISBN 0-7591-0869-2 (pbk. : alk. paper)


1. Ethnology—Methodology. I. Title.

GN345.B36 2006

301'.072—dc22

2005018836

Printed in the United States of America

™ The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI/NISO Z39.48–1992.

Contents

Preface	vii
1. Anthropology and the Social Sciences	1
2. The Foundations of Social Research	28
3. Preparing for Research	69
4. The Literature Search	96
5. Research Design: Experiments and Experimental Thinking	109
6. Sampling	146
7. Sampling Theory	169
8. Nonprobability Sampling and Choosing Informants	186
9. Interviewing: Unstructured and Semistructured	210
10. Structured Interviewing I: Questionnaires	251
11. Structured Interviewing II: Cultural Domain Analysis	299
12. Scales and Scaling	318
13. Participant Observation	342
14. Field Notes: How to Take Them, Code Them, Manage Them	387
15. Direct and Indirect Observation	413
16. Introduction to Qualitative and Quantitative Analysis	451
17. Qualitative Data Analysis I: Text Analysis	463
18. Qualitative Data Analysis II: Models and Matrices	522

19. Univariate Analysis	549
20. Bivariate Analysis: Testing Relations	594
21. Multivariate Analysis	649
Appendix A: Table of Random Numbers	697
Appendix B: Table of Areas under a Normal Curve	700
Appendix C: Student's t Distribution	703
Appendix D: Chi-Square Distribution Table	704
Appendix E: F Tables for the .05 and .01 Levels of Significance	706
Appendix F: Resources for Fieldworkers	710
References	711
Subject Index	771
Author Index	791
About the Author	803

Preface

Since 1988, when I wrote the first edition of this book, I've heard from many colleagues that their departments are offering courses in research methods. This is wonderful. Anthropologists of my generation, trained in the 1950s and 1960s, were hard-pressed to find courses we could take on how to do research. There was something rather mystical about the how-to of fieldwork; it seemed inappropriate to make the experience too methodical.

The mystique is still there. Anthropological fieldwork is fascinating and dangerous. Seriously: Read Nancy Howell's 1990 book on the physical hazards of fieldwork if you think this is a joke. But many anthropologists have found that participant observation loses none of its allure when they collect data systematically and according to a research design. Instead, they learn that having lots of reliable data when they return from fieldwork makes the experience all the more magical.

I wrote this book to make it easier for students to collect reliable data beginning with their first fieldwork experience. We properly challenge one another's explanations for why Hindus don't eat their cattle and why, in some cultures, mothers are more likely than fathers are to abuse their children. That's how knowledge grows. Whatever our theories, though, all of us need data on which to test those theories. The methods for collecting and analyzing data belong to all of us.

What's in This Book

The book begins with a chapter about where I think anthropology fits in the social sciences. With one foot planted squarely in the humanities and the other in the sciences, there has always been a certain tension in the discipline between those who would make anthropology a quantitative science and those whose goal it is to produce documents that convey the richness—indeed, the uniqueness—of human thought and experience.

Students of cultural anthropology and archeology may be asked early in their training to take a stand for qualitative *or* quantitative research. Readers of this textbook will find no support for this pernicious distinction. I lay out my support for positivism in chapter 1, but I also make clear that positivism is not a synonym for quantitative. As you read chapter 1, think about your own position. You don't have to agree with my ideas on epistemological issues to profit from the later chapters on how to select informants, how to choose a sample, how to do questionnaire surveys, how to write and manage field notes, and so on.

Chapter 2 introduces the vocabulary of social research. There's a lot of jargon, but it's the good kind. Important concepts deserve words of their own, and chapter 2 is full of important concepts like reliability, validity, levels of measurement, operationism, and covariation.

Whenever I introduce a new term, like **positivism**, **hermeneutics**, **standard error of the mean**, or whatever, I put it in boldface type. The index shows every example of every boldfaced word. So, if you aren't sure what a **factorial design** is (while you're reading about focus groups in chapter 9, on interviewing), the index will tell you that there are other examples of that piece of jargon in chapter 5 (on experiments), in chapter 10 (on questionnaires), and in chapter 18 (on qualitative analysis).

Chapter 3 is about choosing research topics. We always want our research to be theoretically important, but what does that mean? After you study this chapter, you should know what theory is and how to tell if your research is likely to contribute to theory or not. It may seem incongruous to spend a lot of time talking about theory in a textbook about methods, but it isn't. Theory is about answering research questions . . . and so is method. I don't like the bogus distinction between method and theory, any more than I like the one between qualitative and quantitative. Chapter 3 is also one of several places in the book where I deal with ethics. I don't have a separate chapter on ethics. The topic is important in every phase of research, even in the beginning phase of choosing a problem to study.

Chapter 4 is about searching the literature. Actually, "scouring" is a better word than "searching." In the old days, BC (before computers), you could get away with starting a research paper or a grant proposal with the phrase "little is known about . . ." and filling in the blank. Now, with online databases, you simply can't do that.

Chapter 5 is about research design and the experimental method. You should come away from chapter 5 with a tendency to see the world as a series of natural experiments waiting for your evaluation.

Chapters 6, 7, and 8 are about sampling. Chapter 6 is an introduction to

sampling: why we do it and how samples of individual data and cultural data are different. Chapter 7 is about sampling theory—where we deal with the question “How big should my sample be?” If you’ve had a course in statistics, the concepts in chapter 7 will be familiar to you. If you haven’t had any stats before, read the chapter anyway. Trust me. There is almost no math in chapter 7. The formula for calculating the standard error of the mean has a square root sign. That’s as hard as it gets. If you don’t understand what the standard error is, you have two choices. You can ignore it and concentrate on the concepts that underlie good sampling or you can study chapter 19 on univariate statistics and return to chapter 7 later.

Chapter 8 is about nonprobability sampling and about choosing informants. I introduce the cultural consensus model in this chapter as a way to identify experts in particular cultural domains.

I’ve placed the sampling chapters early in the book because the concepts in these chapters are so important for research design. The validity of research findings depends crucially on measurement; but your ability to generalize from valid findings depends crucially on sampling.

Chapters 9 through 15 are about methods for collecting data. Chapter 9 is titled “Interviewing: Unstructured and Semistructured.” All data gathering in fieldwork boils down to two broad kinds of activities: watching and listening. You can observe people and the environment and you can talk to people and get them to tell you things. Most data collection in anthropology is done by just talking to people. This chapter is about how to do that effectively.

Chapter 10 is devoted entirely to questionnaires—how to write good questions, how to train interviewers, the merits of face-to-face interviews vs. self-administered and telephone interviews, minimizing response effects, and so on.

Chapter 11 is about interviewing methods for cultural domain analysis: pile sorts, triad tests, free listing, frame eliciting, ratings, rankings, and paired comparisons—that is, everything *but* questionnaires.

One topic not covered in chapters 10 and 11 is how to build and use scales to measure concepts. Chapter 12 deals with this topic in depth, including sections on Likert scales and semantic differential scales, two of the most common scaling devices in social research.

Chapter 13 is about participant observation, the core method in cultural anthropology. Participant observation is what produces rapport, and rapport is what makes it possible for anthropologists to do all kinds of otherwise unthinkable intrusive things—watch people bury their dead, accompany fishermen for weeks at a time at sea, ask women how long they breast-feed, go into people’s homes at random times and weigh their food, watch people apply poultices to open sores. . . .

Lone fieldworkers don't have time—even in a year—to interview hundreds and hundreds of people, so our work tends to be less reliable than that of our colleagues in some other disciplines. But participant observation lends validity to our work, and this is a very precious commodity. (More about the difference between reliability and validity in chapter 2.)

Participant observation fieldwork produces field notes—lots of them. Chapter 14 describes how to write and manage field notes.

Chapter 15 is about watching. There are two kinds of watching: the direct, obtrusive kind (standing around with a stopwatch and a note pad) and the indirect, unobtrusive kind (lurking out of sight). Direct observation includes continuous monitoring and spot sampling, and the latter is the method used in time allocation research. Unobtrusive observation poses serious ethical problems, which I treat in some detail in this chapter. One kind of unobtrusive observation poses hardly any ethical problems: research on the physical traces of behavior. You may be surprised at how much you can learn from studying phone bills, marriage contracts, office memos, and other traces of behavior. Your credit rating, after all, is based on other people's evaluation of the traces of your behavior.

Chapters 16 through 21 are about data analysis. Chapter 16 is a general introduction to the fundamentals of analysis. Data do not “speak for themselves.” You have to process data, pore over them, sort them out, and produce an analysis. The canons of science that govern data analysis and the development of explanations apply equally to qualitative and quantitative data.

Chapters 17 and 18 are about the analysis of qualitative data. In chapter 17, I focus on the collection and analysis of texts. There are several traditions of text analysis—hermeneutics, narrative and discourse analysis, grounded theory, content analysis, and schema analysis—some more qualitative, some more quantitative. In chapter 18, I deal with ethnographic decision models and the methods of cognitive anthropology, including the building of folk taxonomies and ethnographic decision-tree modeling.

Chapters 19 through 21 are about the analysis of quantitative data and present the basic concepts of the common statistical techniques used across the social sciences. If you want to become comfortable with statistical analysis, you need more than a basic course; you need a course in regression and applied multivariate analysis and a course (or a lot of hands-on practice) in the use of one of the major statistical packages, like SPSS®, SAS®, and SYSTAT®. Neither the material in this book nor a course in the use of statistical packages is a replacement for taking statistics from professional instructors of that subject. Nevertheless, after working through the materials in chapters 19 through

21, you will be able to use basic statistics to describe your data and you'll be able to take your data to a professional statistical consultant and understand what she or he suggests.

Chapter 19 deals with univariate statistics—that is, statistics that describe a *single variable*, without making any comparisons among variables. Chapters 20 and 21 are discussions of bivariate and multivariate statistics that describe *relationships among variables* and let you test hypotheses about what causes what.

I don't provide exercises at the end of chapters. Instead, throughout the book, you'll find dozens of examples of real research that you can replicate. One of the best ways to learn about research is to repeat someone else's successful project. The best thing about replicating previous research is that *whatever* you find out has to be significant. Whether you corroborate or falsify someone else's findings, you've made a serious contribution to the store of knowledge. If you repeat any of the research projects described in this book, write and tell me about what you found.

What's New in This Edition?

New references have been added throughout the book (the bibliography is about 50% larger than in the last edition) to point students to the literature on the hundreds of methods and techniques covered.

In chapter 1, I've added information on the social science origins of probability theory. I've added several examples of interesting social science variables and units of analysis to chapter 2 and have spelled out the ecological fallacy in a bit more detail. I've added examples (Dordick, Price, Sugita, Edgerton) and have updated some examples in table 3.1. Chapter 4 has been thoroughly updated, including tips on how to search online databases. Some examples of natural experiments were added to chapter 5. In chapter 6, I added examples (Laurent, Miller, Oyuela-Cacedo), and there's a new example on combining probability and nonprobability samples. In chapter 7, I updated the example for the central limit theorem.

Chapter 8, on nonprobability sampling and selecting informants, is much expanded, with more examples and additional coverage of chain referral methods (including snowball sampling), case control sampling, and using consensus analysis to select domain specific informants. In chapter 9, on unstructured and semistructured interviewing, the sections on recording equipment and on voice recognition software (VRS) have been expanded. This may be the last

edition in which I'll talk about tape (rather than digital) recording—though the issue of digital format is hardly settled—and about transcribing machines (rather than about VRS). I've added material in chapter 9 on interviewing with a third party present, on asking threatening questions, and on cued recall to increase the probability of informant accuracy.

In chapter 10, on structured interviewing, I've added a section on computer-based methods, including CASI (computer-assisted self-interviewing), CAPI (computer-assisted personal interviewing), CATI (computer-assisted telephone interviewing), and Internet-based surveys. The chapter has been updated, and there is new material on the social desirability effect, on back translation, on pretesting, on longitudinal surveys, on time budgets, and on mixed methods. In chapter 11, I've added material on free lists and on using paired comparisons to get rank-ordered data. In chapter 12, on scaling, I've added a new example on the semantic differential and a new section on how many choices to offer people in a scaling question.

In chapter 13, on participant observation, I've updated the bibliography and have added new examples of in-home observation (Graham, Sugita), a new example (Wallace) on building awareness, and more material on the importance of learning the native language of the people you're studying. In chapter 14, on taking and managing field notes, I've emphasized the use of computers and have added an example (Gibson) on coding films. Chapter 15, on direct observation, has a new section on ethograms and several new examples, including one (O'Brian) on combining spot sampling and continuous monitoring. Chapter 16, the introduction to general principles of data analysis, is essentially unchanged.

Chapter 17, on text analysis, has been thoroughly updated, with an expanded bibliography, a new section on conversation analysis, and more on how to find themes in text. These new sections owe much to my work with Gery Ryan (see Ryan and Bernard 2000). I've added an example (Paddock) of coding themes in pictures rather than in words and a new example of coding for the Human Relations Area Files (Ember and Ember). I've updated the section on computers and text analysis, but I haven't added instructions on how to use any particular program. I don't do this for *Anthropac*, either, but I discuss the options and point readers to the appropriate websites (and see appendix F). I added more on the native ethnography method in response to Harry Wolcott's cogent critique (1999), and have added a new example for schema analysis.

I continue to add materials on the collection and analysis of visual materials in several parts of the book. For example, chapter 9 has an example of the use of video and photos as cues in an experiment on the accuracy of eyewitness

testimony. There is an example in chapter 14 of coding ethnographic film as text; and there are examples of the use of video in continuous monitoring in chapter 15, along with a description of labanotation, the method used by anthropologists to record physical movements, like dance and nonverbal communication. There is an example of content analysis on a set of films in chapter 17.

However, I don't have a chapter on this vibrant and important set of methods. The field of visual anthropology is developing very quickly with the advent of easy-to-carry, easy-to-use cameras that produce high-quality still and moving images and synchronized sound. Recently, Fadwa El Guindi (2004) published a general text on visual anthropology that covers the whole field: the history of the discipline, ethnographic filmmaking (which she illustrates in detail with her own work), the use of photos as interview probes, the use of film as native ethnography, and the use of photos and film as documentation of culture and culture change.

Chapters 18, 19, and 20 have only minor changes, and, where appropriate, an expanded bibliography. In chapter 21, on multivariate analysis, I've updated some figures in examples, added an extended section on similarity matrices, including tables and a figure, and have rewritten the section on multidimensional scaling with a new example.

Acknowledgments

My debt to colleagues, students, and friends is enormous. Carole Hill, Willett Kempton, William Loker, Kathryn Oths, Aaron Podolefsky, Paul Sabloff, Roger Trent, Douglas Raybeck, and Alvin Wolfe provided helpful criticisms of drafts of earlier editions. Penn Handwerker, Jeffrey Johnson, and Paula Sabloff continue to share ideas with me about teaching research methods. Joseph Bosco, Michael Burton, Michael Chibnik, Art Hansen, William Loker, Kathy Oths, Scott Robinson, Jorge Rocha, Alexander Rodlach, Paula Sabloff, and Christian Sturm were kind enough to report typos and errors in the last edition. In one case, I had calculated incorrectly the numbers of Americans of Chinese, Japanese, and Vietnamese ancestry. Michael Burton's students (Guillermo Narvaez, Allison Fish, Caroline Melly, Neha Vora, and Judith Pajo) went to the census data and corrected my error. I'm very pleased to know that the book is read so carefully and also that students are learning from my mistakes.

Students at the University of Florida have been keen critics of my writing. Domenick Dellino, Michael Evans, Camilla Harshbarger, Fred Hay, Shepherd

Iverson, Christopher McCarty, and David Price were very helpful as I wrote the first edition. Holly Williams, Gery Ryan, Gene Ann Shelley, Barbara Marriott, Kenneth Adams, Susan Stans, Bryan Byrne, and Louis Forline gave me the benefit of their advice for the second edition. Discussions with Nanette Barkey, Lance Gravlee, Harold Green, Scott Hill, David Kennedy, George Mbeh, Isaac Nyamongo, Jorge Rocha, and Kenneth Sturrock helped me with the third edition, as did discussions with Oliver Kortendick, Julia Pauli, and Michael Schnegg at the University of Cologne during 1994–95. And now, for the fourth edition, I thank Stacey Giroux, Mark House, Adam Kiš, Chad Maxwell, Rosalyn Negron, Fatma Soud, Elli Sugita, and Amber Wutich. All have given freely of their time to talk to me about research methods and about how to teach research methods.

Over 40 years of teaching research methods, I have benefited from the many textbooks on the subject in psychology (e.g., Murphy et al. 1937; Kerlinger 1973), sociology (e.g., Goode and Hatt 1952; Lundberg 1964; Nachmias and Nachmias 1976; Babbie 1983), and anthropology (e.g., Johnson 1978; Pelto and Pelto 1978). The scholars whose works most influenced my thinking about research methods were Paul Lazarsfeld (1954, 1982; Lazarsfeld and Rosenberg 1955; Lazarsfeld et al. 1972) and Donald Campbell (1957, 1974, 1975; Campbell and Stanley 1966; Cook and Campbell 1979).

Over those same 40 years, I've profited from discussions about research methods with Michael Agar, Stephen Borgatti, James Boster, Devon Brewer, Ronald Cohen, Roy D'Andrade, William Dressler, Linton Freeman, Sue Freeman, Christina Gladwin, the late Marvin Harris, Penn Handwerker, Jeffrey Johnson, Hartmut Lang, Pertti Pelto, the late Jack Roberts, A. Kimball Romney, Douglas White, Lee Sailer, the late Thomas Schweizer, Susan Weller, and Oswald Werner. Other colleagues who have influenced my thinking about research methods include Ronald Burt, Michael Burton, Carol Ember, Melvin Ember, Eugene Hammel, Allen Johnson, Maxine Margolis, Ronald Rice, Peter Rossi, James Short, Harry Triandis, the late Charles Wagley, Harry Wolcott, and Alvin Wolfe. Most of them knew that they were helping me talk and think through the issues presented in this book, but some may not have, so I take this opportunity to thank them all.

Gery Ryan was my doctoral student, and, as is fitting in such matters, he is now teaching me about methods of text analysis. His influence is particularly important in chapters 17 and 18 in the discussions about coding themes, conversation analysis, and ethnographic decision models.

Time is a gift we all cherish. The first edition of this book was written in 1985–86 during a year of research leave from the University of Florida, for which I thank Charles Sidman, then dean of the College of Liberal Arts and Sciences. I had the opportunity to read widely about research methods and to

begin writing the second edition when I was a guest professor at the Museum of Ethnology in Osaka, Japan, from March to June 1991. My deep appreciation to Kazuko Matsuzawa for that opportunity. A year at the University of Cologne, in 1994–95, as a von Humboldt scholar, gave me the time to continue reading about research methods, across the social and behavioral sciences. Alas, my colleague and host for that year, Thomas Schweizer, died in 1999. The University of Florida granted me a sabbatical to bring out this fourth edition.

In 1987, Pertti Pelto, Lee Sailer, and I taught the first National Science Foundation Summer Institute on Research Methods in Cultural Anthropology—widely known as “methods camp.” Stephen Borgatti joined the team in 1988 (when Sailer left), and the three of us taught together for 8 years, from 1988 to 1995. My intellectual debt to those two colleagues is profound. Pertti Pelto, of course, wrote the pioneering methods text in cultural anthropology (1970), and I’ve long been influenced by his sensible combination of ethnographic and numerical data in field research.

Stephen Borgatti tutored me on the measurement of similarities and dissimilarities and has greatly influenced my thinking about the formal study of emically defined cultural domains. Readers will see many references in this book to Borgatti’s suite of computer programs, called *Anthropac*. That package made it possible for anthropologists to do multidimensional scaling, hierarchical clustering, Likert scaling, Guttman scaling, and other computationally intensive data analysis tasks *in the field*.

The original methods camp, which ended in 1995, was open only to those who already had the Ph.D. In 1996, Jeffrey Johnson founded the NSF Summer Institute for Research Design in Cultural Anthropology. That institute, which continues to this day, is open only to graduate students who are designing their doctoral research. I’ve been privileged to continue to teach at these summer institutes and continue to benefit from collaborating with Johnson and with Susan Weller in teaching young anthropologists the craft of research design.

Penn Handwerker has, for many years, been willing to spend hours on the phone with me, discussing problems of data analysis. My closest colleague, and the one to whom I am most intellectually indebted, is Peter Killworth, with whom I have worked since 1972. Peter is a geophysicist at the University of Southampton and is accustomed to working with data that have been collected by deep-sea current meters, satellite weather scanners, and the like. But he shares my vision of an effective science of humanity, and he has shown an appreciation for the difficulties a naturalist like me encounters in collecting real-life data, in the field, about human behavior and thought. Most importantly, he has helped me see the possibilities for overcoming those difficulties through the application of scientific research practices. The results are never

perfect, but the process of trying is always exhilarating. That's the central lesson of this book, and I hope it comes through.

Mitch Allen commissioned all four editions of this book and has long been a treasured friend and editor. I thank the production staff at Rowman & Littlefield for their thoroughly professional work. It's so important to have really good production people on your side. Speaking of which, anyone who has ever written a book knows the importance of a relentless, take-no-prisoners copy editor. Mine is Carole Bernard. We have a kind of cottage industry: I write, she rips. I am forever grateful.

H. R. B.
August 1, 2005
Gainesville, Florida

1



Anthropology and the Social Sciences

The Craft of Research

This book is about research methods in anthropology—methods for designing research, methods for sampling, methods for collecting data, and methods for analyzing data. And in anthropology, this all has to be done twice, once for qualitative data and once for quantitative data.

No one is expert in all the methods for research. But by the time you get through this book, you'll know about the range of methods used in anthropology and you'll know which kinds of research problems are best addressed by which methods.

Research is a craft. I'm not talking *analogy* here. Research isn't *like* a craft. It *is* a craft. If you know what people have to go through to become skilled carpenters or makers of clothes, you have some idea of what it takes to learn the skills for doing research. It takes practice, practice, and more practice.

Have you ever known a professional seamstress? My wife and I were doing fieldwork in Ixmiquilpan, a small town in the state of Hidalgo, Mexico, in 1962 when we met Florencia. She made dresses for little girls—Communion dresses, mostly. Mothers would bring their girls to Florencia's house. Florencia would look at the girls and say "turn around . . . turn again . . . OK," and that was that. The mother and daughter would leave, and Florencia would start making a dress. No pattern, no elaborate measurement. There would be one fitting to make some adjustments, and that was it.

Carole and I were amazed at Florencia's ability to pick up a scissors and start cutting fabric without a pattern. Then, 2 years later, in 1964, we went to

Greece and met Irini. She made dresses for women on the island of Kalymnos where I did my doctoral fieldwork. Women would bring Irini a catalog or a picture—from Sears or from some Paris fashion show—and Irini would make the dresses. Irini was more cautious than Florencia was. She made lots of measurements and took notes. But there were no patterns. She just looked at her clients, made the measurements, and started cutting fabric.

How do people learn that much? With lots of practice. And that's the way it is with research. Don't expect to do perfect research the first time out. In fact, don't ever expect to do perfect research. Just expect that each time you do a research project, you will bring more and more experience to the effort and that your abilities to gather and analyze data and write up the results will get better and better.

Methods Belong to All of Us

As you go through this book, you'll learn about methods that were developed in other fields as well as methods that were developed in anthropology. In my view, there are no anthropological or sociological or psychological methods. The questions we ask about the human condition may differ across the social sciences, but methods belong to all of us.

Truth is, from the earliest days of the discipline, right up to the present, anthropologists have been prodigious inventors, consumers, and adapters of research methods. Anthropologists developed some of the widely used methods for finding patterns in text, for studying how people use their time, and for learning how people make decisions. Those methods are up for grabs by everyone. The questionnaire survey has been developed mostly by sociologists, but that method is now everyone's. Psychologists make the most consistent use of the experiment, and historians of archives, but anthropologists use and contribute to the improvement of those methods, too.

Anthropologists make the most consistent use of participant observation, but that method turns up in political science, nursing, criminology, and education. The boundaries between the social science disciplines remain strong, but those boundaries are less and less about methods and even less and less about content. Anthropologists are as likely these days as sociologists are to study coming of age in American high schools (Hemmings 2004), how women are socialized to become modern mothers in Greece (Paxon 2004), and alternative medicine in London (Aldridge 2004).

In fact, the differences *within* anthropology and sociology with regard to methods are more important than the differences *between* those disciplines. There is an irreducible difference, for example, between those of us in any of

the social sciences for whom the first principle of inquiry is that reality is constructed uniquely by each person (the **constructivist** view) and those of us who start from the principle that external reality awaits our discovery through a series of increasingly good approximations to the truth (the **positivist** view). There is also an important (but not incompatible) difference between those of us who seek to *understand* people's beliefs and those of us who seek to *explain* what causes those beliefs and action and what those beliefs and actions cause.

Whatever our epistemological differences, though, the actual methods for collecting and analyzing data belong to everyone (Bernard 1993).

Epistemology: Ways of Knowing

The problem with trying to write a book about research methods (besides the fact that there are so *many* of them) is that the word “method” has at least three meanings. At the most general level, it means **epistemology**, or the study of how we know things. At a still-pretty-general level, it's about strategic choices, like whether to do participant observation fieldwork, dig up information from libraries and archives, do a survey, or run an experiment. These are **strategic methods**, which means that they comprise lots of methods at once.

At the specific level, method is about choice of technique—whether to stratify a sample or not, whether to do face-to-face interviews or use the telephone, whether to use a Solomon four-group design or a static-group comparison design in running an experiment, and so on (we'll get to all these things as we go along—experimental designs in chapter 5, sampling in chapters 6, 7, and 8, personal and telephone interviews in chapters 9 and 10, and so on).

When it comes to epistemology, there are several key questions. One is whether you subscribe to the philosophical principles of **rationalism** or **empiricism**. Another is whether you buy the assumptions of the scientific method, often called **positivism** in the social sciences, or favor the competing method, often called **humanism** or **interpretivism**. These are tough questions, with no easy answers. I discuss them in turn.

Rationalism, Empiricism, and Kant

The virtues and dangers of rationalism vs. empiricism have been debated for centuries. Rationalism is the idea that human beings achieve knowledge because of their capacity to reason. From the rationalist perspective, there are a priori truths, which, if we just prepare our minds adequately, will become

evident to us. From this perspective, progress of the human intellect over the centuries has resulted from reason. Many great thinkers, from Plato (428–327 BCE) to Leibnitz (Gottfried Wilhelm Baron von Leibniz, 1646–1716) subscribed to the rationalist principle of knowledge. “We hold these truths to be self-evident” is an example of assuming a priori truths.

The competing epistemology is empiricism. For empiricists, like John Locke (1632–1704), human beings are born **tabula rasa**—with a “clean slate.” What we come to know is the result of our experience written on that slate. David Hume (1711–1776) elaborated the empiricist philosophy of knowledge: We see and hear and taste things, and, as we accumulate experience, we make generalizations. We come, in other words, to understand what is true from what we are exposed to.

This means, Hume held, that we can never be absolutely sure that what we know is true. (By contrast, if we reason our way to a priori truths, we can be certain of whatever knowledge we have gained.) Hume’s brand of **skepticism** is a fundamental principle of modern science. The scientific method, as it’s understood today, involves making incremental improvements in what we know, edging toward truth but never quite getting there—and always being ready to have yesterday’s truths overturned by today’s empirical findings.

Immanuel Kant (1724–1804) proposed a way out, an alternative to either rationalism or empiricism. A priori truths exist, he said, but if we see those truths it’s because of the way our brains are structured. The human mind, said Kant, has a built-in capacity for ordering and organizing sensory experience. This was a powerful idea that led many scholars to look to the human mind itself for clues about how human behavior is ordered.

Noam Chomsky, for example, proposed that any human can learn any language because we have a universal grammar already built into our minds. This would account, he said, for the fact that material from one language can be translated into any other language. A competing theory was proposed by B. F. Skinner, a radical behaviorist. Humans learn their language, Skinner said, the way all animals learn everything, by operant conditioning, or reinforced learning. Babies learn the sounds of their language, for example, because people who speak the language reward babies for making the “right” sounds (see Chomsky 1957, 1969, 1972, 1977; Skinner 1957; Stemmer 1990).

The intellectual clash between empiricism and rationalism creates a dilemma for all social scientists. Empiricism holds that people learn their values and that values are therefore relative. I consider myself an empiricist, but I accept the rationalist idea that there are universal truths about right and wrong.

I’m not in the least interested, for example, in transcending my disgust with, or taking a value-neutral stance about genocide in Germany of the 1940s, or in Cambodia of the 1970s, or in Bosnia and Rwanda of the 1990s, or in Sudan

in 2004–2005. I can never say that the Aztec practice of sacrificing thousands of captured prisoners was just another religious practice that one has to tolerate to be a good cultural relativist. No one has ever found a satisfactory way out of this rationalist-empiricist dilemma. As a practical matter, I recognize that both rationalism and empiricism have contributed to our current understanding of the diversity of human behavior.

Modern social science has its roots in the empiricists of the French and Scottish Enlightenment. The early empiricists of the period, like David Hume, looked outside the human mind, to human behavior and experience, for answers to questions about human differences. They made the idea of a mechanistic science of humanity as plausible as the idea of a mechanistic science of other natural phenomena.

In the rest of this chapter, I outline the assumptions of the scientific method and how they apply to the study of human thought and behavior in the social sciences today.

The Norms of Science

The norms of science are clear. Science is “an objective, logical, and systematic method of analysis of phenomena, devised to permit the accumulation of reliable knowledge” (Lastrucci 1963:6). Three words in Lastrucci’s definition—“objective,” “method,” and “reliable”—are especially important.

1. **Objective.** The idea of truly objective inquiry has long been understood to be a delusion. Scientists do hold, however, that *striving* for objectivity is useful. In practice, this means being explicit about our measurements, so that others can more easily find the errors we make. We constantly try to improve measurement, to make it more precise and more accurate, and we submit our findings to peer review—what Robert Merton called the “organized skepticism” of our colleagues.
2. **Method.** Each scientific discipline has developed a set of techniques for gathering and handling data, but there is, in general, a single scientific method. The method is based on three assumptions: (1) that reality is “out there” to be discovered; (2) that direct observation is the way to discover it; and (3) that material explanations for observable phenomena are always sufficient and metaphysical explanations are never needed. Direct observation can be done with the naked eye or enhanced with various instruments (like microscopes); and human beings can be improved by training as instruments of observation. (I’ll say more about that in chapters 13 and 15 on participant observation and direct observation.)

Metaphysics refers to explanations of phenomena by any nonmaterial force, such as the mind or spirit or a deity—things that, by definition, cannot

be investigated by the methods of science. This does not deny the existence of metaphysical knowledge, but scientific and metaphysical knowledge are quite different. There are time-honored traditions of metaphysical knowledge—knowledge that comes from introspection, self-denial, and spiritual revelation—in cultures across the world.

In fact, science does not reject metaphysical knowledge—though individual scientists may do so—only the use of metaphysics to explain natural phenomena. The great insights about the nature of existence, expressed throughout the ages by poets, theologians, philosophers, historians, and other humanists may one day be understood as biophysical phenomena, but so far, they remain tantalizingly metaphysical.

3. **Reliable.** Something that is true in Detroit is just as true in Vladivostok and Nairobi. Knowledge can be kept secret by nations, but there can never be such a thing as “Venezuelan physics,” “American chemistry,” or “Kenyan geology.”

Not that it hasn't been tried. From around 1935–1965, T. D. Lysenko, with the early help of Josef Stalin, succeeded in gaining absolute power over biology in what was then the Soviet Union. Lysenko developed a Lamarckian theory of genetics, in which human-induced changes in seeds would, he claimed, become inherited. Despite public rebuke from the entire non-Soviet scientific world, Lysenko's “Russian genetics” became official Soviet policy—a policy that nearly ruined agriculture in the Soviet Union and its European satellites well into the 1960s (Joravsky 1970; Soifer 1994; see also Storer 1966, on the norms of science).

The Development of Science: From Democritus to Newton

The scientific method is barely 400 years old, and its systematic application to human thought and behavior is less than half that. Aristotle insisted that knowledge should be based on experience and that conclusions about general cases should be based on the observation of more limited ones. But Aristotle did not advocate disinterested, objective accumulation of reliable knowledge. Moreover, like Aristotle, all scholars until the 17th century relied on metaphysical concepts, like the soul, to explain observable phenomena. Even in the 19th century, biologists still talked about “vital forces” as a way of explaining the existence of life.

Early Greek philosophers, like Democritus (460–370 BCE), who developed the atomic theory of matter, were certainly materialists, but one ancient scholar stands out for the kind of thinking that would eventually divorce sci-

ence from studies of mystical phenomena. In his single surviving work, a poem entitled *On the Nature of the Universe* (1998), Titus Lucretius Carus (98–55 BCE) suggested that everything that existed in the world had to be made of some material substance. Consequently, if the soul and the gods were real, they had to be material, too (see Minadeo 1969). Lucretius' work did not have much impact on the way knowledge was pursued, and even today, his work is little appreciated in the social sciences (but see Harris [1968] for an exception).

Exploration, Printing, and Modern Science

Skip to around 1400, when a series of revolutionary changes began in Europe—some of which are still going on—that transformed Western society and other societies around the world. In 1413, the first Spanish ships began raiding the coast of West Africa, hijacking cargo and capturing slaves from Islamic traders. New tools of navigation (the compass and the sextant) made it possible for adventurous plunderers to go farther and farther from European shores in search of booty.

These breakthroughs were like those in architecture and astronomy by the ancient Mayans and Egyptians. They were based on systematic observation of the natural world, but they were not generated by the social and philosophical enterprise we call science. That required several other revolutions.

Johannes Gutenberg (1397–1468) completed the first edition of the Bible on his newly invented printing press in 1455. (Printing presses had been used earlier in China, Japan, and Korea, but lacked movable type.) By the end of the 15th century, every major city in Europe had a press. Printed books provided a means for the accumulation and distribution of knowledge. Eventually, printing would make organized science possible, but it did not by itself guarantee the objective pursuit of reliable knowledge, any more than the invention of writing had done four millennia before (Eisenstein 1979; Davis 1981).

Martin Luther (1483–1546) was born just 15 years after Gutenberg died. No historical figure is more associated with the Protestant Reformation, which began in 1517, and that event added much to the history of modern science. It challenged the authority of the Roman Catholic Church to be the sole interpreter and disseminator of theological doctrine.

The Protestant affirmation of every person's right to interpret scripture required literacy on the part of everyone, not just the clergy. The printing press made it possible for every family of some means to own and read its own Bible. This promoted widespread literacy, in Europe and later in the

United States. Literacy didn't cause science, but it helped make possible the development of science as an organized activity.

Galileo

The direct philosophical antecedents of modern science came at the end of the 16th century. If I had to pick one single figure on whom to bestow the honor of founding modern science, it would have to be Galileo Galilei (1564–1642). His best-known achievement was his thorough refutation of the Ptolemaic geocentric (Earth-centered) theory of the heavens. But he did more than just insist that scholars *observe* things rather than rely on metaphysical dogma to explain them. He developed the idea of the experiment by causing things to happen (rolling balls down differently inclined planes, for example, to see how fast they go) and measuring the results.

Galileo became professor of mathematics at the University of Padua in 1592 when he was just 28. He developed a new method for making lenses and used the new technology to study the motions of the planets. He concluded that the sun (as Copernicus claimed), not the Earth (as the ancient scholar Ptolemy had claimed) was at the center of the solar system.

This was one more threat to their authority that Roman church leaders didn't need at the time. They already had their hands full, what with break-away factions in the Reformation and other political problems. The church reaffirmed its official support for the Ptolemaic theory, and in 1616 Galileo was ordered not to espouse either his refutation of it or his support for the Copernican heliocentric (sun-centered) theory of the heavens.

Galileo waited 16 years and published the book that established science as an effective method for seeking knowledge. The book's title was *Dialogue Concerning the Two Chief World Systems, Ptolemaic and Copernican*, and it still makes fascinating reading (Galilei 1953 [1632], 1997). Between the direct observational evidence that he had gathered with his telescopes and the mathematical analyses that he developed for making sense of his data, Galileo hardly had to espouse anything. The Ptolemaic theory was simply rendered obsolete.

In 1633, Galileo was convicted by the Inquisition for heresy and disobedience. He was ordered to recant his sinful teachings and was confined to house arrest until his death in 1642. He nearly published *and* perished. For the record, in 1992, Pope John Paul II reversed the Roman Catholic Church's 1616 ban on teaching the Copernican theory and apologized for its condemnation of Galileo. (For more on Galileo, see Drake 1978.)

Bacon and Descartes

Two other figures are often cited as founders of modern scientific thinking: Francis Bacon (1561–1626) and René Descartes (1596–1650). Bacon is known for his emphasis on **induction**, the use of direct observation to confirm ideas and the linking together of observed facts to form theories or explanations of how natural phenomena work. Bacon correctly never told us how to get ideas or how to accomplish the linkage of empirical facts. Those activities remain essentially humanistic—you think hard.

To Bacon goes the dubious honor of being the first “martyr of empiricism.” In March 1626, at the age of 65, Bacon was driving through a rural area north of London. He had an idea that cold might delay the biological process of putrefaction, so he stopped his carriage, bought a hen from a local resident, killed the hen, and stuffed it with snow. Bacon was right—the cold snow did keep the bird from rotting—but he himself caught bronchitis and died a month later (Lea 1980).

Descartes didn’t make any systematic, direct observations—he did neither fieldwork nor experiments—but in his *Discourse on Method* (1960 [1637]) and particularly in his monumental *Meditations* (1993 [1641]), he distinguished between the mind and all external material phenomena—matter—and argued for what is called **dualism** in philosophy, or the independent existence of the physical and the mental world. Descartes also outlined clearly his vision of a universal science of nature based on direct experience and the application of reason—that is, observation and theory. (For more on Descartes’s influence on the development of science, see Schuster 1977, Markie 1986, Hausman and Hausman 1997, and Cottingham 1999.)

Newton

Isaac Newton (1643–1727) pressed the scientific revolution at Cambridge University. He invented calculus and used it to develop celestial mechanics and other areas of physics. Just as important, he devised the **hypothetico-deductive model** of science that combines both **induction** (empirical observation) and **deduction** (reason) into a single, unified method (Toulmin 1980).

In this model, which more accurately reflects how scientists actually conduct their work, it makes no difference where you get an idea: from data, from a conversation with your brother-in-law, or from just plain, hard, reflexive thinking. What matters is whether you can *test* your idea against data in the real world. This model seems rudimentary to us now, but it is of fundamental importance and was quite revolutionary in the late 17th century.

Science, Money, and War

The scientific approach to knowledge was established just as Europe began to experience the growth of industry and the development of large cities. Those cities were filled with uneducated factory laborers. This created a need for increased productivity in agriculture among those not engaged in industrial work.

Optimism for science ran high, as it became obvious that the new method for acquiring knowledge about natural phenomena promised bigger crops, more productive industry, and more successful military campaigns. The organizing mandate for the French Academy of Science in 1666 included a modest proposal to study “the explosive force of gunpowder enclosed (in small amounts) in an iron or very thick copper box” (Easlea 1980:207, 216).

As the potential benefits of science became evident, political support increased across Europe. More scientists were produced; more university posts were created for them to work in. More laboratories were established at academic centers. Journals and learned societies developed as scientists sought more outlets for publishing their work. Sharing knowledge through journals made it easier for scientists to do their own work and to advance through the university ranks. Publishing and sharing knowledge became a material benefit, and the behaviors were soon supported by a value, a norm.

The norm was so strong that European nations at war allowed enemy scientists to cross their borders freely in pursuit of knowledge. In 1780, Reverend Samuel Williams of Harvard University applied for and received a grant from the Massachusetts legislature to observe a total eclipse of the sun predicted for October 27. The perfect spot, he said, was an island off the coast of Massachusetts.

Unfortunately, Williams and his party would have to cross Penobscot Bay. The American Revolutionary War was still on, and the bay was controlled by the British. The speaker of the Massachusetts House of Representatives, John Hancock, wrote a letter to the commander of the British forces, saying “Though we are politically enemies, yet with regard to Science it is presumable we shall not dissent from the practice of civilized people in promoting it” (Rothschild 1981, quoted in Bermant 1982:126). The appeal of one “civilized” person to another worked. Williams got his free passage.

The Development of Social Science: From Newton to Rousseau

It is fashionable these days to say that social science should not imitate physics. As it turns out, physics and social science were developed at about

the same time, and on the same philosophical basis, by two friends, Isaac Newton and John Locke (1632–1704). It would not be until the 19th century that a formal program of applying the scientific method to the study of humanity would be proposed by Auguste Comte, Claude-Henri de Saint-Simon, Adolphe Quételet, and John Stuart Mill (more about these folks in a bit). But Locke understood that the rules of science applied equally to the study of celestial bodies (what Newton was interested in) and to human behavior (what Locke was interested in).

In his *An Essay Concerning Human Understanding* (1996 [1690]), Locke reasoned that since we cannot see everything and since we cannot even record perfectly what we do see, some knowledge will be closer to the truth than will other knowledge. Prediction of the behavior of planets might be more accurate than prediction of human behavior, but both predictions should be based on better and better observation, measurement, and reason (see Nisbet 1980; Woolhouse 1996).

Voltaire, Condorcet, and Rousseau

The legacy of Descartes, Galileo, and Locke was crucial to the 18th-century **Enlightenment** and to the development of social science. Voltaire (François Marie Arouet, 1694–1778) was an outspoken proponent of Newton's nonreligious approach to the study of all natural phenomena, including human behavior (Voltaire 1967 [1738]). In several essays, Voltaire introduced the idea of a science to uncover the laws of history. This was to be a science that could be applied to human affairs and would *enlighten* those who governed so that they might govern better.

Other Enlightenment figures had quite specific ideas about the progress of humanity. Marie Jean de Condorcet (1743–1794) described all of human history in 10 stages, beginning with hunting and gathering, and moving up through pastoralism, agriculture, and several stages of Western states. The ninth stage, he reckoned, began with Descartes and ended with the French Revolution and the founding of the republic. The last stage was the future, reckoned as beginning with the French Revolution.

Jean-Jacques Rousseau (1712–1778), by contrast, believed that humanity had started out in a state of grace, characterized by equality of relations, but that civilization, with its agriculture and commerce, had corrupted humanity and led to slavery, taxation, and other inequalities. Rousseau was not, however, a raving romantic, as is sometimes supposed. He did not advocate that modern people abandon civilization and return to hunt their food in the forests. Rousseau held that the state embodied humanity's efforts, through a **social contract**, to control the evils brought about by civilization. In his clas-

sic work *On The Social Contract*, Rousseau (1988 [1762]) laid out a plan for a state-level society based on equality and agreement between the governed and those who govern.

The Enlightenment philosophers, from Bacon to Rousseau, produced a philosophy that focused on the use of knowledge in service to the improvement of humanity, or, if that weren't possible, at least to the amelioration of its pain. The idea that science and reason could lead humanity toward perfection may seem naive to some people these days, but the ideas of John Locke, Jean Jacques Rousseau, and other Enlightenment figures were built into the writings of Thomas Paine (1737–1809) and Thomas Jefferson (1743–1826), and were incorporated into the rhetoric surrounding rather sophisticated events—like the American and French revolutions. (For more on the history of social science, see Znaniecki 1963 [1952], Olson 1993, McDonald 1994, R. Smith 1997, and Wagner 2001.)

Early Positivism: Quételet, Saint-Simon, Comte

The person most responsible for laying out a program of mechanistic social science was Auguste Comte (1798–1857). In 1824, he wrote: “I believe that I shall succeed in having it recognized . . . that there are laws as well defined for the development of the human species as for the fall of a stone” (quoted in Sarton 1935:10).

Comte could not be bothered with the empirical research required to uncover the Newtonian laws of social evolution that he believed existed. He was content to deduce the social laws and to leave “the verification and development of them to the public” (1875–1877, III:xi; quoted in Harris 1968).

Not so Adolphe Quételet (1796–1874), a Belgian astronomer who turned his skills to both fundamental and applied social research. He developed life expectancy tables for insurance companies and, in his book *A Treatise on Man* (1969 [1842]), he presented statistics on crime and mortality in Europe. The first edition of that book (1835) carried the audacious subtitle “Social Physics,” and, indeed, Quételet extracted some very strong generalizations from his data. He showed that, for the Paris of his day, it was easier to predict the proportion of men of a given age who would be in prison than the proportion of those same men who would die in a given year. “Each age [cohort]” said Quételet, “paid a more uniform and constant tribute to the jail than to the tomb” (1969 [1842]:viii).

Despite Quételet's superior empirical efforts, he did not succeed in building a following around his ideas for social science. But Claude-Henri de Saint-Simon (1760–1825) did, and he was apparently quite a figure. He fought in

the American Revolution, became wealthy in land speculation in France, was imprisoned by Robespierre during the French Revolution, studied science after his release, and went bankrupt living flamboyantly.

Saint-Simon's arrogance must have been something. He proposed that scientists become priests of a new religion that would further the emerging industrial society and would distribute wealth equitably. Saint-Simon's narcissistic ideas were taken up by industrialists after his death in 1825, but the movement broke up in the early 1830s, partly because its treasury was impoverished by paying for some monumental parties (see Durkheim 1958).

Saint-Simon may have been the originator of the **positivist school** of social science, but it was Comte who developed the idea in a series of major books. Comte tried to forge a synthesis of the great ideas of the Enlightenment—the ideas of Kant, Hume, and Voltaire—and he hoped that the new science he envisioned would help to alleviate human suffering. Between 1830 and 1842, Comte published a six-volume work, *The Course of Positive Philosophy*, in which he proposed his famous “law of three stages” through which knowledge developed (see Comte 1853, 1975).

In the first stage of human knowledge, said Comte, phenomena are explained by invoking the existence of capricious gods whose whims can't be predicted by human beings. Comte and his contemporaries proposed that religion itself evolved, beginning with the worship of inanimate objects (fetishism) and moving up through polytheism to monotheism. But any reliance on supernatural forces as explanations for phenomena, said Comte, even a modern belief in a single deity, represented a primitive and ineffectual stage of human knowledge.

Next came the metaphysical stage, in which explanations for observed phenomena are given in terms of “essences,” like the “vital forces” commonly invoked by biologists of the time. The so-called positive stage of human knowledge is reached when people come to rely on empirical data, reason, and the development of scientific laws to explain phenomena. Comte's program of positivism, and his development of a new science he called “sociology,” is contained in his four-volume work *System of Positive Polity*, published between 1875 and 1877.

I share many of the sentiments expressed by the word “positivism,” but I've never liked the word itself. I suppose we're stuck with it. Here is John Stuart Mill (1866) explaining the sentiments of the word to an English-speaking audience: “Whoever regards all events as parts of a constant order, each one being the invariable consequent of some antecedent condition, or combination of conditions, accepts fully the Positive mode of thought” (p. 15) and “All theories in which the ultimate standard of institutions and rules of actions

was the happiness of mankind, and observation and experience the guides . . . are entitled to the name Positive” (p. 69).

Mill thought that the word “positive” was not really suited to English and would have preferred to use “phenomenal” or “experiential” in his translation of Comte. I wish Mill had trusted his gut on that one.

Comte’s Excesses

Comte wanted to call the new positivistic science of humanity “social physiology,” but Saint-Simon had used that term. Comte tried out the term “social physics,” but apparently dropped it when he found that Quételet was using it, too. The term “sociology” became somewhat controversial; language puritans tried for a time to expunge it from the literature on the grounds that it was a bastardization—a mixture of both Latin (*societas*) and Greek (*logo*) roots. Despite the dispute over the name, Comte’s vision of a scientific discipline that both focused on and served society found wide support.

Unfortunately, Comte, like Saint-Simon, had more in mind than just the pursuit of knowledge for the betterment of humankind. Comte envisioned a class of philosophers who, with support from the state, would direct all education. They would advise the government, which would be composed of capitalists “whose dignity and authority,” explained John Stuart Mill, “are to be in the ratio of the degree of generality of their conceptions and operations—bankers at the summit, merchants next, then manufacturers, and agriculturalists at the bottom” (1866:122).

It got worse. Comte proposed his own religion; condemned the study of planets that were not visible to the naked eye; and advocated burning most books except for a hundred or so of the ones that people needed in order to become best educated. “As his thoughts grew more extravagant,” Mill tells us, Comte’s “self-confidence grew more outrageous. The height it ultimately attained must be seen, in his writings, to be believed” (p. 130).

Comte attracted a coterie of admirers who wanted to implement the master’s plans. Mercifully, they are gone (we hope), but for many scholars, the word “positivism” still carries the taint of Comte’s outrageous ego.

The Activist Legacy of Comte’s Positivism

Despite Comte’s excesses, there were three fundamental ideas in his brand of positivism that captured the imagination of many scholars in the 19th century and continue to motivate many social scientists, including me. The first is the idea that the scientific method is the surest way to produce knowledge about the natural world. The second is the idea that scientifically produced

knowledge is effective—it lets us control nature, whether we’re talking about the weather, or disease, or our own fears, or buying habits. And the third is the idea that effective knowledge can be used to improve human lives. As far as I’m concerned, those ideas haven’t lost any of their luster.

Some people are very uncomfortable with this “mastery over nature” metaphor. When all is said and done, though, few people—not even the most outspoken critics of science—would give up the material benefits of science. For example, one of science’s great triumphs over nature is antibiotics. We know that overprescription of those drugs eventually sets the stage for new strains of drug-resistant bacteria, but we also know perfectly well that we’re not going to stop using antibiotics. We’ll rely (we hope) on *more* science to come up with better bacteria fighters.

Air-conditioning is another of science’s triumphs over nature. In Florida, where I live, there is constant criticism of overdevelopment. But try getting middle-class people in my state to give up air-conditioning for even a day in the summer and you’ll find out in a hurry about the weakness of ideology compared to the power of creature comforts. If running air conditioners pollutes the air or uses up fossil fuel, we’ll rely (we hope) on *more* science to solve those problems, too.

Technology and Science

We are accustomed to thinking about the success of the physical and biological sciences, but not about the success of the social sciences. Ask 500 people, as I did in a telephone survey, to list “the major contributions that science has made to humanity” and there is strong consensus: cures for diseases, space exploration, computers, nuclear power, satellite telecommunications, television, automobiles, artificial limbs, and transplant surgery head the list. Not one person—not one—mentioned the discovery of the double helix structure of DNA or Einstein’s theory of relativity.

In other words, the contributions of science are, in the public imagination, technologies—the things that provide the mastery over nature I mentioned.

Ask those same people to list “the major contributions that the social and behavioral sciences have made to humanity” and you get a long silence on the phone, followed by a raggedy list, with no consensus.

I want you to know, right off the bat, that social science is serious business and that it has been a roaring success, contributing mightily to humanity’s global effort to control nature. Everyone in science today, from astronomy to zoology, uses probability theory and the array of statistical tools that have developed from that theory. It is all but forgotten that probability theory was

applied social science right from the start. It was developed in the 17th century by mathematicians Pierre Fermat (1601–1665) and Blaise Pascal (1623–1662) to help people do better in games of chance, and it was well established a century later when two other mathematicians, Daniel Bernoulli (1700–1782) and Jean D’Alambert (1717–1783), debated publicly the pros and cons of large-scale inoculations in Paris against smallpox.

In those days (before Edward Jenner’s breakthrough in 1798), vaccinations against smallpox involved injecting small doses of the live disease. There was a substantial risk of death from the vaccination, but the disease was ravaging cities in Europe and killing people by the thousands. The problem was to assess the probability of dying from smallpox vs. dying from the vaccine. This is one of the earliest uses I have found of social science and probability theory in the making of state policy, but there were soon to be many more. One of them is state lotteries—taxes on people who are bad at math. Another is social security.

In 1889, Otto von Bismarck came up with a pension plan for retired German workers. Based on sound social science data, Bismarck’s minister of finance suggested that 70 would be just the right age for retirement. At that time, the average life expectancy in Germany was closer to 50, and just 30% of children born then could expect to live to 70. Germany lowered the retirement age to 65 in 1916, by which time, life expectancy had edged up a bit—to around 55 (Max-Planck Institute 2002). In 1935, when the Social Security system was signed into law in the United States, Germany’s magic number 65 was adopted as the age of retirement. White children born that year in the United States had an average life expectancy of about 60, and for black children it was only about 52 (SAUS 1947:table 88).

Today, life expectancy in the highly industrialized nations is close to 80—fully 30 years longer than it was 100 years ago—and social science data are being used more than ever in the development of public policy. How much leisure time should we have? What kinds of tax structures are needed to support a medical system that caters to the needs of 80-somethings, when birth rates are low and there are fewer working adults to support the retirement of the elderly?

The success of social science is not all about probability theory and risk assessment. Fundamental breakthroughs by psychologists in understanding the stimulus-response mechanism in humans have made possible the treatment and management of phobias, bringing comfort to untold millions of people. Unfortunately, the same breakthroughs have brought us wildly successful attack ads in politics and millions of adolescents becoming hooked on cigarettes from the likes of Joe Camel. I never said you’d *like* all the successes of social science.

And speaking of great successes that are easy not to like. . . . In 1895, Frederick Winslow Taylor read a paper before the American Society of Mechanical Engineers, entitled “A piece-rate system.” This was the start of scientific management, which brought spectacular gains in productivity and profits—and spectacular gains in worker alienation as well. In 1911, F. B. Gilbreth studied bricklayers. He looked at things like where masons set up their pile of bricks and how far they had to reach to retrieve each brick. From these studies, he made recommendations on how to lessen worker fatigue, increase morale, and raise productivity through conservation of motion.

The method was an instant hit—at least among people who hired bricklayers. Before Gilbreth, the standard in the trade was 120 bricks per hour. After Gilbreth published, the standard reached 350 bricks per hour (Niebel 1982:24). Bricklayers, of course, were less enthusiastic about the new standards.

Just as in the physical and biological sciences, the application of social science knowledge can result in great benefits or great damage to humankind.

Social Science Failures

If the list of successes in the social sciences is long, so is the list of failures. School busing to achieve racial integration was based on scientific findings in a report by James Coleman (1966). Those findings were achieved in the best tradition of careful scholarship. They just happened to be wrong because the scientists involved in the study couldn’t anticipate “white flight”—a phenomenon in which Whites abandoned cities for suburbs, taking much of the urban tax base with them and driving the inner cities further into poverty.

On the other hand, the list of failures in the physical and biological sciences is quite impressive. In the Middle Ages, alchemists tried everything they could to turn lead into gold. They had lots of people investing in them, but it just didn’t work. Cold fusion is still a dream that attracts a few hardy souls. And no one who saw the explosion of the space shuttle *Challenger* on live television in January 1986 will ever forget it.

There are some really important lessons from all this. (1) Science isn’t perfect but it isn’t going away because it’s just too successful at doing what people everywhere want it to do. (2) The sciences of human thought and human behavior are much, much more powerful than most people understand them to be. (3) The power of social science, like that of the physical and biological sciences, comes from the same source: the scientific method in which ideas, based on hunches or on formal theories, are put forward, tested publicly, and replaced by ideas that produce better results. And (4) social science knowl-

edge, like that of any science, can be used to enhance our lives or to degrade them.

The Varieties of Positivism

These days, positivism is often linked to support for whatever power relations happen to be in place. It's an astonishing turnabout, because historically, positivism was linked to social activism. In *The Subjection of Women* (1869), John Stuart Mill advocated full equality for women, and Adolphe Quételet, the Belgian astronomer whose study of demography and criminology carried the audacious title *Social Physics* (1969 [1835]), was a committed social reformer.

The legacy of positivism as a vehicle for social activism is clear in Jane Addams's work with destitute immigrants at Chicago's Hull House (1926), in Sidney and Beatrice Webb's attack on the abuses of the British medical system (1910), in Charles Booth's account of the living conditions of the poor in London (1902), and in Florence Nightingale's (1871) assessment of death rates in maternity hospitals. (See McDonald [1993] for an extended account of Nightingale's long-ignored research.)

The central position of positivism is that experience is the foundation of knowledge. We record what we experience—what we see others do, what we hear others say, what we feel others feel. The quality of the recording, then, becomes the key to knowledge. Can we, in fact, record what others do, say, and feel? Yes, of course we can. Are there pitfalls in doing so? Yes, of course there are. To some social researchers, these pitfalls are evidence of natural limits to a science of humanity; to others, like me, they are a challenge to extend the current limits by improving measurement. The fact that knowledge is tentative is something we all learn to live with.

Later Positivism: The Vienna Circle

Positivism has taken some interesting turns. Ernst Mach (1838–1916), an Austrian physicist, took an arch-empiricist stance further than even Hume might have done himself: If you could not verify something, Mach insisted, then you should question its existence. If you can't see it, it isn't there. This stance led Mach to reject the atomic theory of physics because, at the time, atoms could not be seen.

Discussion of Mach's ideas was the basis of a seminar group that met in Vienna and Berlin during the 1920s and 1930s. The group, composed of math-

emancipators, philosophers, and physicists, came to be known as the **Vienna Circle** of logical positivists. They were also known as logical empiricists, and when social scientists today discuss positivism, it is often this particular brand that they have in mind (see Mach 1976).

The term **logical empiricism** better reflects the philosophy of knowledge of the members of the Vienna Circle than does **logical positivism**. Unfortunately, Herbert Feigl and Albert Blumberg used “logical positivism” in the title of their 1931 article in the *Journal of Philosophy* in which they laid out the program of their movement, and the name “positivism” stuck—again (Smith 1986).

The fundamental principles of the Vienna Circle were that knowledge is based on experience and that metaphysical explanations of phenomena were incompatible with science. Science and philosophy, they said, should attempt to answer only scientifically answerable questions. A question like “Was Mozart or Brahms the better composer?” can only be addressed by metaphysics and should be left to artists.

In fact, the logical positivists of the Vienna Circle did not see art—painting, sculpture, poetry, music, literature, and literary criticism—as being in conflict with science. The arts, they said, allow people to express personal visions and emotions and are legitimate unto themselves. Since poets do not claim that their ideas are testable expressions of reality, their ideas can be judged on their own merits as either evocative and insightful, or not. Therefore, any source of wisdom (like poetry) that generates ideas, and science, which tests ideas, are mutually supportive and compatible (Feigl 1980).

I find this eminently sensible. Sometimes, when I read a really great line of poetry, like Robert Frost’s line from *The Mending Wall*, “Good fences make good neighbors,” I think “How could I *test* that? Do good fences *always* make good neighbors?” When shepherders fenced off grazing lands across the western United States in the 19th century, keeping cattle out of certain regions, it started range wars.

Listen to what Frost had to say about this in the same poem: “Before I built a wall I’d ask to know/ What I was walling in or walling out./ And to whom I was like to give offence.” The way I see it, the search for understanding is a human activity, no matter who does it and no matter what epistemological assumptions they follow.

Understanding begins with questions and with ideas about how things work. When do fences make good neighbors? Why do women earn less, on average, for the same work as men in most industrialized countries? Why is Barbados’s birth rate falling faster than Saudi Arabia’s? Why is there such a high rate of alcoholism on Native American reservations? Why do nation states, from Italy to Kenya, almost universally discourage people from maintaining minority

languages? Why do public housing programs often wind up as slums? If advertising can get children hooked on cigarettes, why is public service advertising so ineffective in lowering the incidence of high-risk sex among adolescents?

Instrumental Positivism

The practice that many researchers today love to hate, however, is neither the positivism of Auguste Comte nor that of the Vienna Circle. It is, instead, what Christopher Bryant (1985:137) calls “instrumental positivism.”

In his 1929 presidential address to the American Sociological Society, William F. Ogburn laid out the rules. In turning sociology into a science, he said, “it will be necessary to crush out emotion.” Further, “it will be desirable to taboo ethics and values (except in choosing problems); and it will be inevitable that we shall have to spend most of our time doing hard, dull, tedious, and routine tasks” (Ogburn 1930:10). Eventually, he said, there would be no need for a separate field of statistics because “all sociologists will be statisticians” (p. 6).

The Reaction against Positivism

That kind of rhetoric just begged to be reviled. In *The Counter-Revolution of Science*, Friedrich von Hayek (1952) laid out the case against the possibility of what Ogburn imagined would be a science of humanity. In the social sciences, Hayek said, we deal with mental phenomena, not with material facts. The data of the social sciences, Hayek insisted, are not susceptible to treatment as if they were data from the natural world. To pretend that they are is what he called “scientism.”

Furthermore, said Hayek, scientism is more than just foolish. It is evil. The ideas of Comte and of Marx, said Hayek, gave people the false idea that governments and economies could be managed scientifically and this, he concluded, had encouraged the development of the communism and totalitarianism that seemed to be sweeping the world when he was writing in the 1950s (Hayek 1952:110, 206).

I have long appreciated Hayek’s impassioned and articulate caution about the need to protect liberty, but he was wrong about positivism, and even about scientism. Science did not cause Nazi or Soviet tyranny any more than religion caused the tyranny of the Crusades or the burning of witches in 17th-century Salem, Massachusetts. Tyrants of every generation have used any means,

including any convenient epistemology or cosmology, to justify and further their despicable behavior. Whether tyrants seek to justify their power by claiming that they speak to the gods or to scientists, the awful result is the same. But the *explanation* for tyranny is surely neither religion nor science.

It is also apparent that an effective science of human behavior exists, no matter whether it's called positivism or scientism or human engineering or anything else. However distasteful it may be to some, John Stuart Mill's simple formula for a science applied to the study of human phenomena has been very successful in helping us understand (and control) human thought and behavior. Whether we like the outcomes is a matter of conscience, but no amount of moralizing diminishes the fact of success.

Today's truths are tomorrow's rubbish, in anthropology just as in physics, and no epistemological tradition has a patent on interesting questions or on good ideas about the answers to such questions. Several competing traditions offer alternatives to positivism in the social sciences. These include humanism, hermeneutics, and phenomenology.

Humanism

Humanism is an intellectual tradition that traces its roots to Protagoras' (485–410 BC) famous dictum that "Man is the measure of all things," which means that truth is not absolute but is decided by individual human judgment. Humanism has been historically at odds with the philosophy of knowledge represented by science.

Ferdinand C. S. Schiller (1864–1937), for example, was a leader of the European humanist revolt against positivism. He argued that since the method and contents of science are the products of human thought, reality and truth could not be "out there" to be found, as positivists assume, but must be made up by human beings (Schiller 1969 [1903]).

Wilhelm Dilthey (1833–1911) was another leader of the revolt against positivism in the social sciences. He argued that the methods of the physical sciences, although undeniably effective for the study of inanimate objects, were inappropriate for the study of human beings. There were, he insisted, two distinct kinds of sciences: the **Geisteswissenschaften** and the **Naturwissenschaften**—that is, the human sciences and the natural sciences. Human beings live in a web of meanings that they spin themselves. To study humans, he argued, we need to understand those meanings (Dilthey 1985 [1883]. For more on Dilthey's work, see Hodges 1952.)

Humanists, then, do not deny the effectiveness of science for the study of nonhuman objects, but emphasize the uniqueness of humanity and the need for a different (that is, nonscientific) method for studying human beings. Simi-

larly, scientists do not deny the inherent value of humanistic knowledge. To explore whether King Lear is to be pitied as a pathetic leader or admired as a successful one is an exercise in seeking humanistic knowledge. The answer to the question cannot possibly be achieved by the scientific method.

In any event, finding *the* answer to the question is not important. Carefully *examining* the question of Lear, however, and producing many possible answers, leads to insight about the human condition. And that *is* important.

Just as there are many competing definitions of positivism, so there are for humanism as well. Humanism is often used as a synonym for humanitarian or compassionate values and a commitment to the amelioration of suffering. The problem is that died-in-the-wool positivists can also be committed to humanitarian values. Counting the dead *accurately* in so-called collateral damage in war, for example, is a very good way to preserve outrage. We need more, not less, science, lots and lots more, and more humanistically informed science, to contribute more to the amelioration of suffering and the weakening of false ideologies—racism, sexism, ethnic nationalism—in the world.

Humanism sometimes means a commitment to subjectivity—that is, to using our own feelings, values, and beliefs to achieve insight into the nature of human experience. In fact, **trained subjectivity** is the foundation of clinical disciplines, like psychology, as well as the foundation of participant observation ethnography. It isn't something apart from social science. (See Berg and Smith [1985] for a review of clinical methods in social research.)

Humanism sometimes means an appreciation of the unique in human experience. Writing a story about the thrill or the pain of giving birth, about surviving hand-to-hand combat, about living with AIDS, about winning or losing a long struggle with illness—or writing someone else's story for them, as ethnographers often do—are not activities *opposed* to a natural science of experience. They *are* the activities of a natural science of experience.

Hermeneutics

The ancient Greek god, Hermes, had the job of delivering and interpreting for humans the messages of the other gods. From this came the Greek word *hermeneus*, or interpreter, and from that comes our word **hermeneutics**, the continual interpretation and reinterpretation of texts.

The idea that texts have meaning and that interpretation can get at that meaning is nothing new. Literacy in ancient Greece and Rome involved the ability to discuss and interpret texts. The Talmud—a series of interpretations of the Five Books of Moses compiled over several hundred years beginning in the second century CE—is a massive hermeneutic exercise. And the great

concordances and exegetical commentaries on the New Testament are a form of hermeneutics.

In biblical hermeneutics, it is assumed that the Bible contains truths and that human beings can extract those truths through careful study and constant interpretation and reinterpretation. In the United States, we treat the Constitution as a sacred document that contains timeless truths, and we interpret and reinterpret the document to see how those truths should play out over time. The same Constitution has, at various times, permitted or forbade slavery, permitted or forbade universal voting rights, and so on.

The hermeneutic tradition has come into the social sciences with the close and careful study of all free-flowing texts. In anthropology, the texts may be myths or folk tales. The hermeneutic approach would stress that: (1) The myths contain some underlying meaning, at least for the people who tell the myths; and (2) It is our job to discover that meaning, knowing that the meaning can change over time and can also be different for subgroups within a society. Think, for example, of the stories taught in U.S. schools about Columbus's voyages. The meaning of those stories may be quite different for Navajos, urban African Americans, Chicanos, and Americans of northern and central European descent.

The hermeneutic approach—the discovery of the meaning of texts through constant interpretation and reinterpretation—is easily extended to the study of any body of texts: sets of political speeches, letters from soldiers in battle to their families at home, transcriptions of doctor-patient interactions. The idea that culture is “an assemblage of texts” is the basis for the interpretive anthropology of Clifford Geertz (1973). And Paul Ricoeur, arguing that action, like the written word, has meaning to actors, extended the hermeneutic approach even to free-flowing behavior itself (1981, 1991). In fact, portable camcorders make it easy to capture the natural behavior of people dancing, singing, interacting over meals, telling stories, and participating in events. In chapter 17 on text analysis, we'll look at how anthropologists apply the hermeneutic model to the study of culture.

Phenomenology

Like positivism, **phenomenology** is a philosophy of knowledge that emphasizes direct observation of phenomena. Unlike positivists, however, phenomenologists seek to *sense* reality and to describe it in words, rather than numbers—words that reflect consciousness and perception. Phenomenology is part of the humanistic tradition that emphasizes the common experience of all human beings and our ability to relate to the feelings of others (see Veatch 1969).

The philosophical foundations of phenomenology were developed by Edmund Husserl (1859–1938), who argued that the scientific method, appropriate for the study of physical phenomena, was inappropriate for the study of human thought and action (see Husserl 1964 [1907], 1999). Husserl's ideas were elaborated by Alfred Schutz, and Schutz's version of phenomenology has had a major impact in social science, particularly in psychology but also in anthropology.

When you study molecules, Schutz said, you don't have to worry about what the world "means" to the molecules (1962:59). But when you try to understand the reality of a human being, it's a different matter entirely. The only way to understand social reality, said Schutz, was through the meanings that people give to that reality. In a phenomenological study, the researcher tries to see reality through another person's eyes.

Phenomenologists try to produce convincing descriptions of what they experience rather than explanations and causes. Good ethnography—a narrative that describes a culture or a part of a culture—is usually good phenomenology, and there is still no substitute for a good story, well told, especially if you're trying to make people understand how the people you've studied think and feel about their lives. (For more on phenomenology, see Moran 2000, Sokolowski 2000, Zahavi 2003, and Elliott 2005.)

About Numbers and Words: The Qualitative/Quantitative Split

The split between the positivistic approach and the interpretive-phenomenological approach pervades the human sciences. In psychology and social psychology, most *research* is in the positivistic tradition, while much *clinical* work is in the interpretivist tradition because, as its practitioners cogently point out, it works. In sociology, there is a growing tradition of interpretive research, but most sociology is done from the positivist perspective.

In anthropology, the situation is a bit more complicated. Most anthropological *data collection* is done by fieldworkers who go out and stay out, watch and listen, take notes, and bring it all home. This makes anthropology a thoroughly empirical enterprise. But much of anthropological *data analysis* is done in the interpretivist tradition, and some empirical anthropologists reject the positivist epistemological tradition, while other empirical anthropologists (like me) identify with that tradition.

Notice in the last two paragraphs the use of words like "approach," "perspective," "tradition," and "epistemology." Not once did I say that "research in X is mostly quantitative" or that "research in Y is mostly qualitative." That's because a commitment to an interpretivist or a positivist epistemology is independent of any commitment to, or skill for, quantification. Searching

the Bible for statistical evidence to support the subjugation of women doesn't turn the enterprise into science.

By the same token, at the early stages of its development, any science relies primarily on qualitative data. Long before the application of mathematics to describe the dynamics of avian flight, qualitative, fieldworking ornithologists did systematic observation and recorded (in words) data about such things as wing movements, perching stance, hovering patterns, and so on. Qualitative description is a kind of measurement, an integral part of the complex whole that comprises scientific research.

As sciences mature, they come inevitably to depend more and more on quantitative data and on quantitative tests of qualitatively described relations. But this never, ever lessens the need for or the importance of qualitative research in any science.

For example, qualitative research might lead us to say that "most of the land in Popotlán is controlled by a minority." Later, quantitative research might result in our saying "76% of the land in Popotlán is controlled by 14% of the inhabitants." The first statement is not wrong, but its sentiment is confirmed and made stronger by the second statement. If it turned out that "54% of the land is controlled by 41% of the inhabitants," then the first part of the qualitative statement would still be true—more than 50% of the land is owned by less than 50% of the people, so most of the land is, indeed controlled by a minority—but the sentiment of the qualitative assertion would be rendered weak by the quantitative observations.

For anthropologists whose work is in the humanistic, phenomenological tradition, quantification is inappropriate. And for those whose work is in the positivist tradition, it is important to remember that numbers do not automatically make any inquiry scientific. In chapter 17, I'll discuss how texts—including words and pictures—can be collected and analyzed by scholars who identify with either the positivist or the interpretivist tradition.

In the rest of this book, you'll read about methods for describing individuals and groups of people. Some of those methods involve library work, some involve controlled experiments, and some involve fieldwork. Some methods result in words, others in numbers. Never use the distinction between quantitative and qualitative as cover for talking about the difference between science and humanism. Lots of scientists do their work without numbers, and many scientists whose work is highly quantitative consider themselves humanists.

Ethics and Social Science

The biggest problem in conducting a science of human behavior is not selecting the right sample size or making the right measurement. It's doing

those things ethically, so you can live with the consequences of your actions. I'm not exaggerating about this. Ethics is part of method in science, just as it is in medicine, business, or any other part of life. For while philosophers discuss the fine points of whether a true science of human behavior is really possible, effective social science is being done all the time, and with rather spectacular, if sometimes disturbing, success.

In the mid-19th century, when Quételet and Comte were laying down the program for a science of human affairs, no one could predict the outcome of elections, or help people through crippling phobias with behavior modification, or engineer the increased consumption of a particular brand of cigarettes. We may question the wisdom of engineering cigarette purchases in the first place, but the fact remains, we *can* do these things, we *are* doing these things, and we're getting better and better at it all the time.

It hardly needs to be pointed out that the increasing effectiveness of science over the past few centuries has also given human beings the ability to cause greater environmental degradation, to spread tyranny, and even to cause the ultimate, planetary catastrophe through nuclear war. This makes a science of humanity even more important now than it has ever been before.

Consider this: Marketers in a midwestern city, using the latest supercomputers, found that if someone bought disposable diapers at 5 P.M., the next thing he or she was likely to buy was a six-pack of beer. So they set up a display of chips next to the disposable diapers and increased snack sales by 17% (Wilke 1992). At the time, 15 years ago, that was a breakthrough in the monitoring of consumer behavior. Today, every time you buy something on the Internet or download a computer program or a piece of music, you leave a trail of information about yourself and your consumer preferences. By tracking your purchases over time, and by sharing information about your buying behavior across websites, market researchers develop ads that are targeted just for you.

We need to turn our skills in the production of such effective knowledge to solving the problems of hunger, disease, poverty, war, environmental pollution, family and ethnic violence, and racism, among others. Social scientists, including anthropologists, can play an important role in social change by predicting the consequences of ethically mandated programs and by refuting false notions (such as various forms of racism) that are inherent in most popular ethical systems. This has been a hallmark of anthropology since Franz Boas's devastating critique, nearly a century ago, of racial theories about why some ethnic minorities in the United States were taller and healthier than others.

Don't get me wrong here. The people who discovered that fact about the six packs and the diapers were good scientists, as are the people who design all those automated data-collection mechanisms for monitoring your behavior on the Internet. I'm not calling for rules to make all those scientists work on

problems that I think are important. Scientists choose to study the things that industry and government pay for, and those things change from country to country and from time to time in the same country. Science has to earn its support by producing useful knowledge. What “useful” means, however, changes from time to time even in the same society, depending on all sorts of historical circumstances.

Suppose we agreed that “useful” means to save lives. AIDS is a terrible disease, but three times as many people died in motor vehicle accidents in 2002 as died of AIDS (about 44,000 and 14,000 respectively). Should we spend three times more money teaching safe driving than we do teaching safe sex?

I think the answer is pretty clear. In a democracy, researchers and activists want the freedom to put their skills and energies to work on what they think is important. Fortunately, that’s just how it is, and, personally, I hope it stays just that way.

2



The Foundations of Social Research

The Language of Social Research

This chapter is about the fundamental concepts of social research: **variables, measurement, validity, reliability, cause and effect, and theory.** When you finish this chapter, you should understand the crucial role of measurement in science and the mutually supportive roles of data and ideas in the development of theory.

You should also have a new skill: You should be able to **operationalize** any complex human phenomenon, like “machismo” or “anomie” or “alienation” or “acculturation.” You should, in other words, be able to reduce any complex variable to a set of measurable traits.

By the end of this chapter, though, you should also become very critical of your new ability at operationalizing. Just because you *can* make up measurements doesn’t guarantee that they’ll be useful or meaningful. The better you get at concocting clever measurements for complex things, the more critical you’ll become of your own concoctions and those of others.

Variables

A **variable** is something that can take more than one value. The *values can be words or numbers.* If you ask a woman how old she was at her first pregnancy, the answer will be a number (16 or 40, or whatever), but if you ask her about her religion, the answer will be a word (“Muslim” or “Methodist”).

Social research is based on defining variables, looking for associations among them, and trying to understand whether—and how—variation in one

thing causes variation in another. Some common variables that you'll find in social research are age, sex, ethnicity, race, education, income, marital status, and occupation.

A few of the hundreds of variables you'll see in anthropological research include number of children by each of several wives in a polygynous household, distance from a clinic or a market or a source of clean water, blood pressure, and level of support for various causes (the distribution of clean needles to drug addicts, the new farmer's co-op, rebels fighting in Eritrea, etc.).

Variables Have Dimensions

Variables can be **unidimensional** or **multidimensional**. The distance from Boston to Denver can be expressed in driving time or in miles, but no matter how you measure it, distance is expressed as a straight line and straight lines are one dimensional. You can see this in figure 2.1.

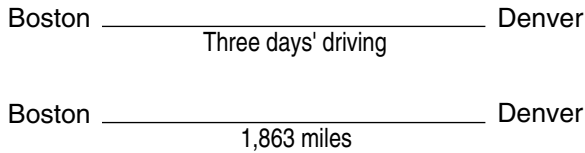


Figure 2.1. Two ways to measure distance.

If we add Miami, we have three distances: Boston-Miami, Boston-Denver, Denver-Miami. One dimension isn't enough to express the relation among three cities. We have to use two dimensions. Look at figure 2.2.

The two dimensions in figure 2.2 are up-down and right-left, or North-South and East-West. If we add Nairobi to the exercise, we'd either have to add a third dimension (straight through the paper at a slight downward angle from Denver), or do what Gerardus Mercator (1512–1594) did to force a three-dimensional object (the Earth) into a two-dimensional picture. Mercator was able to project a sphere in two dimensions, but at the cost of distortion at the edges. This is why, on a map of the world, Greenland (an island of 840,000 square miles), looks the same size as China (a land mass of about 3.7 million square miles).

Height, weight, birth order, age, and marital status are unidimensional variables and are relatively easy to measure. By contrast, political orientation (being conservative or liberal) is multidimensional and is, therefore, a lot more difficult to measure. We often talk about political orientation as if it were unidimensional, with people lying somewhere along a line between strictly con-

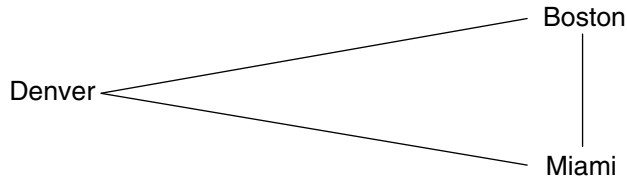


Figure 2.2. Three points create two dimensions.

servative and strictly liberal. But if you think about it, people can be liberal about some dimensions of life and conservative about others. For example, you might agree strongly with the statement that “men and women should get equal pay for equal work” and also with the statement that “the war in Iraq is necessary to defend freedom in America.” These statements test political orientation about domestic economic policy and foreign policy—two of the many dimensions of political orientation.

Even something as seemingly straightforward as income is multidimensional. To measure the annual income of retired Americans in Florida, for example, you have to account for social security benefits, private pension funds, gifts from children and other kin, gambling winnings, tax credits, interest on savings, wages paid entirely in cash (including tips), food stamps. . . .

And don’t think it’s easier in out-of-the-way communities around the world. If you think it’s tough assessing the amount that a waitress earns from tips, try assessing the amount a Haitian family gets from people who are working in Miami and sending money home.

In chapter 12, after we look at questionnaire design, I’ll discuss the building of scales and how to test for the unidimensionality of variables.

Simplifying Variables: Race and Gender

In the United States, at least, race is treated (by academics as well as by people in general) as a **dichotomous variable**, with two values: black and white. This makes race easy to measure and, in fact, we’ve learned a lot by making the measurement simple. For example, any man in the United States who is labeled “black” is about five times more likely to be the victim of homicide than is any man labeled “white.” This is down from a ratio of nearly eight-to-one in 1991. Black babies are about two-and-a-half times more likely to die in infancy than are white babies, and people labeled “black” are two-and-a-half times more likely as people labeled “white” to be poor (which meant \$18,392 for a family of four in 2002) (SAUS 2004–2005, tables 100, 297, 685, 686).

Still, we know that there are gradations of skin color besides black and white, so it's reasonable to ask whether people who are *more* black are *more* likely to be a victim of homicide, to die in infancy, to be poor, etc. Around 1970, medical researchers began to find a relation in the United States between darkness of skin color and blood pressure among people labeled "Blacks" (see Boyle 1970; Harburg et al. 1978). The darker the skin, the higher blood pressure was likely to be.

Later, researchers began to find that education and social class were more important predictors of high blood pressure among Blacks than was darkness of skin color (see Keil et al. 1977, 1981). This meant that darker-skinned people were more likely to be the victims of discrimination and, as a consequence, uneducated and poor. Poverty causes stress and poor diet, both of which are direct causes of high blood pressure.

But suppose we treated skin color as the continuous variable it really is rather than as a dichotomous variable? Clarence Gravlee (2002b) did this in his study of race and blood pressure in Puerto Rico. He measured skin color in two ways. First, he showed people a line with nine numbers on it and asked them to rate themselves from light to dark by telling him which number best described their skin color. Then he measured the color of people's inner arm with a photospectrometer. The first measure is *emic* (what people think, themselves, about their color) and the second is *etic* (an objective, external measurement that doesn't depend on what people think).

Now, *etic* skin color—the amount of melanin that people have in their skin, as measured by a photospectrometer—by itself doesn't account for variation in blood pressure. But the *difference* between *etic* skin color and what people *say* their color is is strongly associated with people's blood pressure (Gravlee 2002b:182). The relationship between these variables is anything but simple. Poor people who rate themselves as having darker skin than they really have are likely to have higher blood pressure. For middle-class people, it's the other way around: They are likely to have lower blood pressure when they rate their skin color as darker than it really is. The puzzle requires a lot more work, but this much is clear: Variation in blood pressure is not caused by melanin (Gravlee and Dressler 2005).

It may not be possible for everyone who uses skin color as an independent variable to measure it with a photospectrometer (the gadgets are very expensive), but if we did this, we could assess whether white schoolteachers react more negatively to darker-skinned black children than they do to lighter-skinned black children, and if so, by how much. This would help us account for some of the variation in black children's school scores as a function of teacher reaction to skin color. This, in turn, would show *how* skin color leads

to discrimination in education, *how* discrimination in education leads to poverty and *how* all this leads to lowered life expectancy.

We already know that Whites live longer than Blacks do. Making skin color a continuous variable would help us learn how racism actually works, not just its consequences.

If the benefits of such research are attractive, though, consider the risks. Racists might claim that our findings support their despicable ideas about the genetic inferiority of African Americans. Life insurance companies might start charging premiums based on amount of skin pigmentation. Even if the Supreme Court ruled against this practice, how many people would be hurt before the matter was adjudicated? As you can see, every research question has an ethical component.

Gender is another dichotomous variable (male and female) that is more complex than it seems. We usually measure gender according to the presence of male or female sexual characteristics. Then we look at the relation between the presence of those characteristics and things like income, level of education, amount of labor migration, attitudes to various social issues, aptitude for math, success in certain jobs, and so on.

But if you think about it, we're not interested in whether differences in human anatomy predict any of these things. What we really want to know is how being *more* male or *more* female (socially and psychologically) predicts attitudes about social issues, success in various jobs, and many other things—like the ability to secure agricultural credit, the ability to cope with widowhood, or health status in old age.

Sandra Bem (1974, 1979) developed a scale called the BSRI (Bem Sex Role Inventory) to measure sex-role identity. The scale consists of 60 words or phrases: 20 that represent what Americans in the early 1970s generally thought of as masculine traits (like independent and assertive); 20 that represented generally accepted feminine traits (like affectionate and sympathetic); and 20 that represented generally accepted gender-neutral traits (like tactful and happy). Respondents rate themselves on a scale of 1 to 7 on how much they think each trait applies to them. Depending on your score, you are either “sex typed” (displaying stereotyped feminine traits or masculine traits) or androgynous (getting a high score on both feminine and masculine traits) or undifferentiated (getting a low score on both feminine and masculine traits).

As you can imagine, the BSRI has gotten plenty of criticism over the years, and, to be sure, what people in the United States think of as typically masculine or feminine traits has changed in the last three decades, but the BSRI has been used in hundreds of studies across many Western societies and in some non-Western societies as well. For example, Sundvik and Lindeman (1993) applied the BSRI to 257 managers (159 men and 98 women) of a government-

controlled transportation company in Finland. Each of the managers had rated a subordinate on 30 dimensions—things like the ability to get along with others, independence in getting the job done, willingness to implement innovations, and so on. The sex-typed female managers (the women who scored high on femaleness, according to the BSRI) rated their male subordinates more favorably than they rated their female subordinates. Similarly, the sex-typed male managers rated their female subordinates more favorably than they rated their male subordinates.

The bottom line, according to Sundvik and Lindeman: “Among persons whose self-concepts are formed on the basis of gender, both the queen bee and the king ape syndromes are alive and well” (1993:8). Sex-typed managers discriminate against subordinates of the same sex.

Of course, traits thought to be masculine in one culture might be thought of as feminine in another. Aggressiveness is a trait widely viewed across many cultures to be desirable for men and boys and undesirable for women and girls. In Zimbabwe, however, 488 schoolteachers, half of whom were men, gave this trait their lowest desirability rating of the 20 masculine items in the BSRI (Wilson et al. 1990).

In Japan, Katsurada and Sugihara (1999) found that all 20 masculine traits in the BSRI were culturally appropriate, but that three of the classically 20 feminine traits in the scale (“sensitive to the needs of others,” “understanding,” and “loyal”) were inappropriate. (Loyalty, for example, is seen as a highly desirable trait for everyone in Japan, so it can’t be used in a test to distinguish between men and women.) Based on tests with 300 college students, Katsurada and Sugihara recommend substituting “conscientious,” “tactful,” and “happy” in the list of feminine adjectives when the BSRI is used in Japan. (For a version of the BSRI tested for use in Mexico, see Lara-Cantú and Navarro-Arias 1987. For a version of the BSRI for use in China, see Qin and Yianjie 2003.) After 25 years of research with the BSRI, we’ve learned a lot about the differences between men and women.

One thing we’ve learned is that those differences are much more complex than a biological dichotomy would make them appear to be. We’ve also learned that gender role differences are even more complex than Bem imagined. Choi and Fuqua (2003) looked at 23 validation studies of the BSRI and found that Bem’s inventory doesn’t fully capture the complexity of masculinity and femininity. But that just means that we’re learning more with each generation of researchers—exactly what we expect from a cumulative science. (For more on measuring gender across cultures using the PAQ and the BSRI, see Sugihara and Warner 1999, Auster and Ohm 2000, Sugihara and Katsurada 2000, Zhang et al. 2001, and Norvilitis and Reid 2002.)

Dependent and Independent Variables

Beginning in the 1840s, breakthroughs in sociology and anthropology produced insight into the impact of economic and political forces on demography. One practical result of all this work was life insurance. The way life insurance works is that you bet the company that you'll die within 365 days. You answer a few questions (How old are you? Do you smoke? What do you do for a living? Do you fly a small plane?), and the company sets the odds—say, your \$235 against the company's promise to pay your heirs \$100,000 if you win the bet and die within 365 days. But if you *lose* the bet and stay alive, they keep your \$235, and next year you go through all this again, except that now the odds are raised against you to say, your \$300 against the company's promise to pay your heirs a lot of money.

For insurance companies to turn a profit, they have to win more bets than they lose. They can make mistakes at the individual level, but in the aggregate (that is, averaging over all people) they have to predict longevity from things they can measure.

Longevity, then, is the **dependent variable**, because it *depends on* sex, education, occupation, etc. These latter are called **independent variables** because they are logically prior to, and therefore independent of, the dependent variable of longevity. How long you live doesn't have any effect on your sex. In our earlier example, blood pressure was the dependent variable. There is no way skin color depends on a person's blood pressure.

It's not always easy to tell whether a variable is independent or dependent. Does high female infant mortality among Amazonian tribal people depend on high levels of warfare, or is it the other way around? Does high income depend on having a lot of land, or vice versa? Do inner-city adolescent girls get pregnant because they are poor, or . . . ? Does the need for litigation stimulate the production of attorneys, or . . . ?

Failure to understand which of two variables depends on the other is the source of endless shenanigans. One of my teachers, Oscar Lewis (1961, 1965), described what he called a "culture of poverty" among slum dwellers in cities around the world. People who live in a culture of poverty, said Lewis, are not very future oriented. This plays out, he said, in their shopping for food every day and in never buying large economy sizes of anything. Lewis's point was that truly poor people can't invest in soap futures by buying large boxes of it. He saw a low level of expressed orientation toward the future, then, as the dependent variable and poverty as the independent variable.

Many people interpreted Lewis's work as meaning exactly the opposite: that poverty is caused by a low level of future orientation. According to this topsy-turvy, victim-blaming reasoning, if poor people everywhere would just

learn to save their money and invest in the future, then they could break the poverty cycle. Such reasoning may serve to create pointless programs to teach poor people how to save money they don't have, but it doesn't do much else.

In rural West Virginia, for example, there is a lot of teen pregnancy and many adolescents drop out of high school. Since the 1960s, according to Bickel et al. (1997), state policymakers in West Virginia have blamed these behaviors on the culture of poverty. The behaviors that state policymakers want so much to change, however, are caused by the continuing deterioration of economic and social conditions in rural communities. No amount of educating poor people about their bad habits will change the material circumstances that cause the so-called culture of poverty.

This **educational model of social change** is a lesson in confusion about dependent and independent variables. The model is based on the attractive idea that, since the last thing that happens before an action is a thought, if you want to create better actions then you need to create better thoughts. In other words, if you want to change people's behavior, you have to change how they think: Teach women in India the value of small families so they'll use birth control to prevent unwanted pregnancies; teach Kenyans why it's important to use bed nets to prevent malaria; teach farmers across the world the importance of washing their hands after handling manure and before preparing or eating food.

The educational model is the basis for one of the world's biggest industries—social change and development—but the model is mostly ineffective because behavioral change (the supposed dependent variable) doesn't usually depend on education (the supposed independent variable). In fact, across the developing world, when women have access to well-paying jobs outside the home, they tend to lower their fertility. Once that happens, they encourage their daughters to stay in school longer. Education doesn't just cause jobs to happen. Instead, jobs for women in one generation cause education in the next. (I'll have more to say on fertility control and the educational model of behavioral change in chapter 3, when I discuss the role of theory in the development of research questions.)

Measurement and Concepts

Variables are measured by their **indicators**, and indicators are defined by their **values**. Some variables, and their indicators, are easily observed and measured. Others are more conceptual. The difference is important.

Consider the variables race and gender again. If skin color can take one of two values (black or white), then to *measure* race you simply look at a person

and decide which value to record. If you use secondary sexual characteristics as an indicator of gender, then to *measure* gender you look at a person and decide whether they are female or male.

In other words, *measurement is deciding which value to record*. That decision is prone to error. Some people whom you classify as white or black might be classified as black or white by another observer. And gender is even worse. Many people, both men and women, have ambiguous secondary sexual characteristics and many women wear what were once considered to be men's clothes. Is Pat a man's name or a woman's? What about Chris? Leslie? Any of these indicators may lead you into making the wrong measurement—marking down a man or boy as a woman or girl, or vice versa.

Improving measurement in science means lowering the probability of and the amount of error. Light-skinned African Americans who cease to identify themselves ethnically as black persons count on those errors for what they hope will be upward economic mobility. Dark-skinned “Whites,” like some Americans of Mediterranean descent, sometimes complain that they are being “mistaken for” Blacks and discriminated against.

Race and gender are concepts or constructs. We have to make them up to study them. All variables are concepts, but some concepts, like height and weight, are easy to measure, while other concepts like religious intensity, jealousy, compassion, willingness to accept new agricultural technologies, and tolerance for foreign fieldwork are complex and difficult to measure.

We are led to defining constructs by our experience: Some people just seem more religiously intense than others, more jealous than others, more tolerant of foreign fieldwork than others, etc. We verify our intuition about conceptual variables by measuring them, or by measuring their results.

Suppose you put an ad in the paper that says: “Roommate wanted. Easy-going, nonsmoker preferred.” When people answer the ad you can look at their fingers and smell their clothes to see if they smoke. But you have to ask people a series of indicator *questions* to gauge their easy-goingness.

Similarly, if you are doing fieldwork in a Peruvian highland village, and you want to predict who among the villagers is predisposed to migrate to the coast in search of work, you will want to measure that predisposition with a series of indicators. In this case, the indicators can be answers to questions (“Have you ever thought about migrating?”). Or they might be observable facts (Does a person have a close relative who has already migrated?). Or they might be a combination of these.

It may be easier to measure some concepts than others, but the fact is, all measurement is difficult. People have worked for centuries to develop good instruments for measuring things like temperature. And if it's difficult to measure temperature (a concept, after all, backed up by time-tested theories), how

do you measure future orientation or machismo? Measuring variables like these is one of our biggest challenges because these variables are mostly what we're interested in.

One of the most famous variables in all of social science is “socioeconomic status” (SES). Measuring it is no easy task. You can use income as one indicator, but there are many wealthy people who have low SES (the so-called *nouveau riche*), and many relatively low-income people who have high SES (think of those down-at-the-heels nobles in England who have to open their castles to tourists to make ends meet).

You can add “level of education” to income as an indicator, but that still won't be enough in most societies of the world to get at something as multidimensional as SES. You can add occupation, father's occupation, number of generations in a community, and so on, depending on the group you are studying, and you still might wind up dissatisfied with the result if your measure fails to predict some dependent variable of interest.

And, as you saw with the Bem androgyny scale earlier, indicators of any concept may vary from culture to culture. This doesn't mean that measurement is impossible. *It means that you have to test (and, if necessary, adapt) every measure of every variable in every new culture where you want to use it.*

Conceptual and Operational Definitions

While most of the interesting variables in social science are concepts, some of our most important concepts are not variables. The concept of “positivism” is *not* a variable, but the concept of “philosophies of science” *is* a variable, and positivism is one member of the list of those philosophies. The concept of “love” is not a variable, but the concept of “being in love or not” *is* one. The concept of “culture” is not a variable, but the concept of “belonging to a particular culture” *is* one. The concept of “attitude” is not a variable, but the concept of “supporting the idea that clitoridectomy is a violation of fundamental human rights” implies an attitude variable with at least two attributes, support and nonsupport.

Conceptual Definitions

There are two ways to define variables—conceptually and operationally. **Conceptual definitions** are abstractions, articulated in words, that facilitate understanding. They are the sort of definitions we see in dictionaries, and we

use them in everyday conversation to tell people what we mean by some term or phrase. **Operational definitions** consist of a set of instructions on how to measure a variable that has been conceptually defined.

Suppose I tell you that “Alice and Fred just moved to a spacious house.” Nice concept. You ask: “What do you mean by ‘spacious’?” and I say: “You know, big rooms, high ceilings.”

If that isn’t enough for you, we’ll have to move from a conceptual definition of “spacious” to an operational one. We’ll have to agree on what to measure: Do we count the screened-in porch and the garage or just the interior living space? Do we count the square footage or the cubic footage? That is, do we get a measure of the living surface, or some measure of the “feeling of spaciousness” that comes from high ceilings? Do we measure the square footage of open space before or after the furniture and appliances go in? If we had to agree on things like this for every concept, ordinary human discourse would come to a grinding halt.

Science is not ordinary human discourse, however, and this, in my view, is the most important difference between the humanistic and the scientific (positivistic) approaches to social science. Humanistic researchers seek to maintain the essential feel of human discourse. Positivists focus more on specific measurement. I do not see these two styles as inimical to one another, but as complementary.

To get a feel for how complementary the two styles can be, ask some 50 year olds and some 20 year olds—men and women of both ages—to tell you how old you have to be in order to be middle aged. You’ll see immediately how volatile the conceptual definition of “middle age” is. If you ask people about what it *means* to “be middle aged,” you’ll get plenty of material for an interesting paper on the subject. If you want to *measure* the differences between men and women and between older and younger people on this variable, you’ll have to do more than just ask them. Figure 2.3 shows an instrument for measuring this variable.

1 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100

Here is a line that represents age. Obviously, a person 1 year of age is a baby, and a person 100 years of age is old. Put a mark on the line where you think middle age begins and another mark where you think middle age ends.

Figure 2.3. An instrument for measuring what people think “middle age” means.

Many concepts that we use in anthropology have volatile definitions: “power,” “social class,” “machismo,” “alienation,” “willingness to change,” and “fear of retribution.” If we are to talk sensibly about such things, we need clear, **intersubjective** definitions of them. In other words,

although there can be no objective definition of middle age, we can at least agree on what we mean by “middle age” for a particular study and on how to measure the concept.

Complex variables are conceptually defined by reducing them to a series of simpler variables. Saying that “the people in this village are highly acculturated” can be interpreted in many ways. But if you state clearly that you include “being bilingual,” “working in the national economy,” and “going to school” in your conceptual definition of acculturation, then at least others will understand what you’re talking about when you say that people are “highly acculturated.”

Similarly, “machismo” might be characterized by “a general feeling of male superiority,” accompanied by “insecure behavior in relationships with women.” Intelligence might be conceptually defined as “the ability to think in abstractions and to generalize from cases.” These definitions have something important in common: They have no external reality against which to test their truth value.

Conceptual definitions are at their most powerful when they are linked together to build theories that explain research results. When the United Nations was founded in 1945, the hope was that trade between industrialized and nonindustrialized countries of the world would result in economic development for everyone. The economies of the developed countries would expand and the benefits of an expanding economy would be seen in the underdeveloped countries. A decade later, it was obvious that this wasn’t what was happening. The rich countries were getting richer and the poor countries were getting poorer.

Raul Prebisch, an Argentinian economist who worked at the UN, argued that under colonialism, rich countries were importing raw materials from poor countries to produce manufactured goods and that poor countries had come to depend economically on the rich countries. Prebisch’s “dependency theory” links the concept of “control of capital” with those of “mutual security” and “economic dependency,” and the linkage helps explain why economic development often results in some groups winding up with less access to capital than they had before a development program (Prebisch 1984, 1994).

Conceptual definitions are at their weakest in the conduct of research itself, because concepts have no empirical basis—we have to make them up to study them.

There is nothing wrong with this. There are three things one wants to do in any science: (1) describe a phenomenon of interest; (2) explain what causes it; and (3) predict what it causes. The existence of a conceptual variable is inferred from what it predicts—how well it makes theoretical sense out of a lot of data.

The Concept of Intelligence

The classic example of a conceptual variable is intelligence. Intelligence is anything we say it is. There is no way to tell whether it is: (1) the ability to think in abstractions and to generalize from cases; (2) the ability to remember long strings of unconnected facts; or (3) the ability to recite all of Shakespeare from memory. In the last analysis, the value of the concept of intelligence is that it allows us to predict, *with varying success*, things like job success, grade-point average, likelihood of having healthy children, and likelihood of being arrested for a felony.

The key to understanding the last statement is the phrase “with varying success.” It is by now well known that measures of intelligence are culture bound; the standard U.S. intelligence tests are biased in favor of Whites and against African Americans because of differences in access to education and differences in life experiences. Further afield, intelligence tests that are designed for Americans may not have any meaning at all to people in radically different cultures.

There is a famous, perhaps apocryphal, story about some American researchers who were determined to develop a culture-free intelligence test based on manipulating and matching shapes and colors. With an interpreter along for guidance, they administered the test to a group of Bushmen in the Kalahari Desert of South Africa. The first Bushman they tested listened politely to the instructions about matching the colors and shapes and then excused himself.

He returned in a few minutes with half a dozen others, and they began an animated discussion about the test. The researchers asked the interpreter to explain that each man had to take the test himself. The Bushmen responded by saying how silly that was; they solve problems together, and they would solve this one, too. So, although the content of the test might have been culture free, the testing procedure itself was not.

This critique of intelligence *testing* in no way lessens the importance or usefulness of the *concept* of intelligence. The concept is useful, in certain contexts, because its measurement allows us to predict other things we want to know. And it is to actual measurement that we now turn.

Operational Definitions

Conceptual definitions are limited because, while they point us toward measurement, they don't really give us any recipe for measurement. Without measurement, we cannot make useful comparisons. We cannot tell whether Spaniards are more flamboyant than the British, or whether Catholicism is more

authoritarian than Buddhism. We cannot evaluate the level of anger in an urban community over perceived abuses by the police of their authority, or compare the level of that anger to the anger found in another community in another city.

Operational definitions specify exactly what you have to do to measure something that has been defined conceptually. Here are four examples of operational definitions:

1. Intelligence: Take the Wechsler Adults Intelligence Scale (WAIS) and administer it to a person. Count up the score. Whatever score the person gets is his or her intelligence.
2. Machismo: Ask a man if he approves of women working outside the home, assuming the family doesn't need the money; if he says "no," then give him a score of 1, and if he says "yes," score him 0. Ask him if he thinks women and men should have the same sexual freedom before marriage; if he says "no," score 1 and score 0 for "yes." Ask him if a man should be punished for killing his wife and her lover; if he says "no," score 1; score 0 for "yes." Add the scores. A man who scores 3 has more machismo than a man who scores 2, and a man who scores 2 has more machismo than a man who scores 1.
3. Tribal identity: Ask American Indians if they speak the language of their ancestors fluently. If "yes," score 1. If "no," score 0. Ask them if they attend at least one tribal pow-wow each year. Score 1 for "yes," and 0 for "no." Ask them eight other questions of this type, and give them a score of 1 for each answer that signifies self-identification with their tribal heritage. Anyone who scores at least 6 out of 10 is an "identifier." Five or less is a "rejecter" of tribal heritage or identity.
4. Support for trade barriers against China: Ask workers in a textile factory to complete the Support of Trade Barriers against China Scale. Add the four parts of the scale together to produce a single score. Record that score.

These definitions sound pretty boring, but think about this: If you and I use the same definitions for variables, *and if we stick to those definitions in making measurements*, then our data are strictly comparable:

We can tell if children in city A have higher intelligence scores than do children in city B.

We can tell if older men in Huehuetenango have higher machismo scores than do younger men in that same village.

We can tell if people in tribe A have higher cultural identity scores than do people in tribe B.

We can tell whether the average scores indicating level of support for trade barriers against China is greater among workers in the factory you studied than it is among workers in the factory I studied.

I find the ability to make such comparisons exciting, and not at all boring. But did you notice that I *never* said anything in those comparisons about eth-

nic identity per se, or intelligence per se, or machismo per se, or support for trade barriers per se. In each case, all I said was that we could tell if the *scores* were bigger or smaller.

What's So Good about Operationism?

Operational definitions are *strictly limited to the content of the operations specified*. That's why I also didn't say anything about whether it was a good idea or a bad one to make any of these measurements or comparisons. *If the content of an operational definition is bad, then so are all conclusions you draw from using it to measure something.*

This is *not* an argument against operationism in science. Just the opposite. Operationism is the best way to expose bad measurement. By defining measurements operationally, we can tell if one measurement is better than another. If the operational measurement of, say, machismo, seems silly or offensive, it may be because the concept is not very useful to begin with. No amount of measurement or operationism bails out bad concepts. The act of trying, though, usually *exposes* bad concepts and helps you jettison them.

Adhering to bad measurements is bad science and can have some bad consequences for people. In the 1960s, I was a consultant on a project that was supposed to help Chicano high schoolers develop good career aspirations. Studies had been conducted in which Chicano and Anglo high schoolers were asked what they wanted to be when they reached 30 years of age. Chicanos expressed, on average, a lower occupational aspiration than did Anglos. This led some social scientists to advise policymakers that Chicano youth needed reinforcement of career aspirations at home. (There's that educational model again.)

Contrary to survey findings, ethnographic research showed that Chicano parents had very high aspirations for their children. The parents were frustrated by two things: (1) despair over the cost of sending their children to college; and (2) high school counselors who systematically encouraged Chicana girls to become housewives and Chicano boys to learn a trade or go into the armed services.

The presumed relation between the dependent variable (level of career aspiration) and the independent variable (level of aspiration by parents for the careers of their children) was backward. The parents' level of career aspiration for their children didn't cause the children to have low aspirations. The children were driven to low aspirations by structural features of their environment. The parents of those children reflected this reality *in order*—they said explicitly to interviewers who bothered to ask—not to give their children false hopes.

The operational definition of the variable “parents’ career aspirations for their children” was useless. Here’s the operational definition that should have been used in the study of Chicano parents’ aspirations for their children’s careers:

Go to the homes of the respondents. Using the native language of the respondents (Spanish or English as the case may be), talk to parents about what they want their high school-age children to be doing in 10 years. Explore each answer in depth and find out why parents give each answer.

Ask specifically if the parents are telling you what they think their children *will* be doing or what they *want* their children to be doing. If parents hesitate, say: “Suppose nothing stood in the way of your [son] [daughter] becoming anything they wanted to be. What would you like them to be doing ten years from now?”

Write down what the parents say and code it for the following possible scores: 1 = unambivalently in favor of children going into high-status occupations; 2 = ambivalent about children going into high-status occupations; 3 = unambivalently in favor of children going into low- or middle-status occupations.

Use Stricker’s (1988) occupation scale to decide whether the occupations selected by parents as fitting for their children are high, middle, or low status. Be sure to take and keep notes on what parents say are the reasons for their selections of occupations.

Notice that taking an ethnographic—a so-called qualitative—approach did not stop us from being operational.

Operationism is often crude, but that, too, can be a strength. Robert Wuthnow (1976) operationalized the concept of religiosity in 43 countries using UNESCO data on the number of books published in those countries and the fraction of those books classified as religious literature. Now *that’s* crude. Still, Wuthnow’s measure of “average religiosity” correlates with seven out of eight indicators of modernity. For example, the higher the literacy rate in 1952, the lower the religiosity in 1972.

I have no idea what that means, but I think following up Wuthnow’s work with more refined measurements—to test hypotheses about the societal conditions that support or weaken religiosity—is a lot more exciting than dismissing it because it was so audaciously crude.

The Problem with Operationism

Strict operationism creates a knotty philosophical problem. We make up concepts and measurement turns these abstractions into reality. Since there are many ways to measure the same abstraction, the reality of any concept hinges

on the device you use to measure it. So, sea temperature is different if you measure it from a satellite (you get an answer based on radiation) or with a thermometer (you get an answer based on a column of mercury). Intelligence is different if you measure it with a Stanford-Binet test, or the Wechsler scales. If you ask a person in any of the industrialized nations “How old are you?” or “How many birthdays have you had?” you will probably retrieve the same number. But the very concept of age in the two cases is different because different instruments (queries are instruments) were used to measure it.

This principle was articulated in 1927 by Percy Bridgman in *The Logic of Modern Physics*, and has become the source of an enduring controversy. The bottom line on strict operational definitions is this: No matter how much you insist that intelligence is really more than what is measured by an intelligence test, that’s all it can ever be. Whatever you think intelligence is, it is exactly and only what you measure with an intelligence test and nothing more.

If you don’t like the results of your measurement, then build a better test, where “better” means that the outcomes are more useful in building theory, in making predictions, and in engineering behavior.

I see no reason to waffle about this, or to look for philosophically palatable ways to soften the principle here. The science that emerges from a strict operational approach to understanding variables is much too powerful to water down with backpedaling. It is obvious that “future orientation” is more than my asking someone “Do you buy large or small boxes of soap?” The problem is, *you* might not include that question in your interview of the same respondent unless I specify that I asked that question in that particular way.

Operational definitions permit scientists to talk to one another using the same language. They permit replication of research and the accumulation of knowledge about issues of importance. The Attitudes Toward Women Scale (AWS) was developed by Janet Spence and Robert Helmreich in 1972. Through 1995, the scale had been applied 71 times to samples of American undergraduate students (Twenge 1997).

Some of the items on the AWS seem pretty old-fashioned today. For example, in one item, people are asked how much they agree or disagree with the idea that “women should worry less about their rights and more about becoming good wives and mothers.” You probably wouldn’t use that item if you were building an attitudes-toward-women scale today, but keeping the original, 1972 AWS intact over all this time lets us track attitudes toward women over time.

The results are enlightening. Attitudes toward women have, as you’d guess, become consistently more liberal/feminist over time, but men’s support for women’s rights have lagged behind women’s support by about 15 years: Men’s average score on the AWS in 1990 was about the same as women’s

average score in 1975 (Twenge 1997). And these data, remember, reflect the attitudes of college students—the quarter of the population whom we expect to be at the vanguard of social change. As the AWS gets more and more out of date, it gets used less frequently, but each time it does get used, it provides another set of data about how attitudes toward women have changed over time and across cultures. (For an assessment of the AWS, see Loo and Thorpe 1998.)

Levels of Measurement

Whenever you define a variable operationally, you do so at some **level of measurement**. Most social scientists recognize the following four levels of measurement, in ascending order: nominal, ordinal, interval, and ratio. The general principle in research is: Always use the highest level of measurement that you can. (This principle will be clear by the time you get through the next couple of pages.)

Nominal Variables

A variable is something that can take more than one value. The values of a **nominal variable** comprise a list of names (name is *nomen* in Latin). You can list religions, occupations, and ethnic groups; and you can also list fruits, emotions, body parts, things to do on the weekend, baseball teams, rock stars . . . the list of things you can list is endless.

Think of nominal variables as *questions, the answers to which tell you nothing about degree or amount*. What's your name? In what country were you born? Are you healthy? On the whole, do you think the economy is in good shape? Is Mexico in Latin America? Is Bangladesh a poor country? Is Switzerland a rich country?

The following survey item is an operationalization of the nominal variable called “religious affiliation”:

- 26a. Do you identify with any religion? (check one)
- Yes No
- If you checked “yes,” then please answer question 26b.
- 26b. What is your religion? (check one):
- Protestant
- Catholic
- Jewish
- Moslem
- Other religion
- No religion

This operationalization of the variable “religious affiliation” has two important characteristics: It is *exhaustive* and *mutually exclusive*. The famous “other” category in nominal variables makes the list exhaustive—that is, all possible categories have been named in the list—and the instruction to “check one” makes the list mutually exclusive. (More on this in chapter 10 when we discuss questionnaire design.)

“Mutually exclusive” means that things can’t belong to more than one category of a nominal variable at a time. We assume, for example, that people who say they are Catholic generally don’t say they are Moslem. I say “generally” because life is complicated and variables that seem mutually exclusive may not be. Some citizens of Lebanon have one Catholic and one Moslem parent and may think of themselves as both Moslem and Catholic.

Most people think of themselves as either male or female, but not everyone does. The prevalence of transsexuals in human populations is not known precisely, but worldwide, it is likely to be between one in ten thousand and one in a hundred thousand for male-to-female transsexuals (biological males whose gender identity is female) and between one in a hundred thousand and one in four hundred thousand for female-to-male transsexuals (Cohen-Kettenis and Gooren 1999).

Most people think of themselves as a member of one so-called race or another, but more and more people think of themselves as belonging to two or more races. In 2000, the U.S. Census offered people the opportunity to check off more than one race from six choices: White, Black or African American, American Indian or Alaska Native, Asian, Native Hawaiian and other Pacific islander, and some other race. Nearly seven million people (2.4% of the 281 million in the United States in 2000) checked more than one of the six options (Grieco and Cassidy 2001).

And when it comes to ethnicity, the requirement for mutual exclusivity is just hopeless. There are Chicano African Americans, Chinese Cuban Americans, Filipino Cherokees, and so on. This just reflects the complexity of real life, but it does make analyzing data more complicated since each *combination of attributes* has to be treated as a separate category of the variable “ethnicity” or collapsed into one of the larger categories. More about this in chapters 19 and 20, when we get to data analysis.

Occupation is a nominal variable, but lots of people have more than one occupation. People can be peasant farmers and makers of fireworks displays for festivals; they can be herbalists and jewelers; or they can be pediatric oncology nurses and antique car salespeople at the same time. A list of occupations is a measuring instrument at the nominal level: You hold each person up against the list and see which occupation(s) he or she has (have).

Nominal measurement—naming things—is **qualitative measurement**.

When you assign the numeral 1 to men and 2 to women, all you are doing is substituting one kind of name for another. Calling men 1 and women 2 does not make the variable quantitative. The number 2 happens to be twice as big as the number 1, but this fact is meaningless with nominal variables. You can't add up all the 1s and 2s and calculate the "average sex" any more than you can add up all the telephone numbers in the Chicago phone book and get the average phone number.

Assigning numbers to things makes it easier to do certain kinds of statistical analysis on qualitative data (more on this in chapter 17), but it doesn't turn qualitative variables into quantitative ones.

Ordinal Variables

Like nominal-level variables, **ordinal variables** are generally exhaustive and mutually exclusive, but they have one additional property: Their values can be rank ordered. Any variable measured as high, medium, or low, like socioeconomic class, is ordinal. The three classes are, in theory, mutually exclusive and exhaustive. In addition, a person who is labeled "middle class" is lower in the social class hierarchy than someone labeled "high class" and higher in the same hierarchy than someone labeled "lower class." What ordinal variables do not tell us is *how much* more.

Scales of opinion—like the familiar "strongly agree," "agree," "neutral," "disagree," "strongly disagree" found on so many surveys—are ordinal measures. They measure an internal state, agreement, in terms of *less* and *more*, but not in terms of *how much* more.

This is the most important characteristic of ordinal measures: There is no way to tell how far apart the attributes are from one another. A person who is middle class might be twice as wealthy and three times as educated as a person who is lower class. Or they might be three times as wealthy and four times as educated. A person who "agrees strongly" with a statement may agree twice as much as someone who says they "agree"—or eight times as much, or half again as much. There is no way to tell.

Interval and Ratio Variables

Interval variables have all the properties of nominal and ordinal variables. They are an exhaustive and mutually exclusive list of attributes, and the attributes have a rank-order structure. They have one additional property, as well: The distances between the attributes are meaningful. Interval variables, then, involve true **quantitative measurement**.

The difference between 30°C and 40°C is the same 10° as the difference

between 70° and 80°, and the difference between an IQ score of 90 and 100 is (assumed to be) the same as the difference between one of 130 and 140. On the other hand, 80 Fahrenheit is not twice as hot as 40, and a person who has an IQ of 150 is not 50% smarter than a person who has an IQ of 100.

Ratio variables are interval variables that have a true **zero point**—that is, a 0 that measures the absence of the phenomenon being measured. The Kelvin scale of temperature has a true zero: It identifies the absence of molecular movement, or heat.

The consequence of a true zero point is that measures have ratio properties. A person who is 40 years old is 10 years older than a person who is 30, and a person who is 20 is 10 years older than a person who is 10. The 10-year intervals between the attributes (years are the attributes of age) are identical. That much is true of an interval variable. In addition, however, a person who is 20 is twice as old as a person who is 10; and a person who is 40 is twice as old as a person who is 20. These, then, are true ratios.

While temperature (in Fahrenheit or Celsius) and IQ are nonratio interval variables, most interval-level variables in the social sciences are also ratio variables. In fact, it has become common practice in the social sciences to refer to ratio-level variables as interval variables and vice versa. This is not technically pure, but the confusion of the terms “interval” and “ratio” doesn’t cause much real damage.

Some examples of ratio variables include: age, number of times a person has changed residence, income in dollars or other currency, years married, years spent migrating, population size, distance in meters from a house to a well, number of hospital beds per million population, number of months since last employment, number of kilograms of fish caught per week, number of hours per week spent in food preparation activities. Number of years of education is usually treated as a ratio variable, even though a year of grade school is hardly worth the same as a year of graduate school.

In general, **concepts** (like alienation, political orientation, level of assimilation) are measured at the ordinal level. People get a high score for being “very assimilated,” a low score for being “unassimilated,” and a medium score for being “somewhat assimilated.” When a concept variable like intelligence is measured at the interval level, it is likely to be the focus of a lot of controversy regarding the validity of the measuring instrument.

Concrete **observables**—things you can actually see—are often measured at the interval level. But not always. Observing whether a woman has a job outside her home is nominal, qualitative measurement based on direct observation.

A Rule about Measurement

Remember this rule: Always measure things at the highest level of measurement possible. Don't measure things at the ordinal level if you can measure them as ratio variables.

If you really want to know the price that people paid for their homes, then ask the price. Don't ask them whether they paid "less than a million pesos, between a million and five million, or more than five million." If you really want to know how much education people have had, ask them how many years they went to school. Don't ask: "Have you completed grade school, high school, some college, four years of college?"

This kind of packaging just throws away information by turning interval-level variables into ordinal ones. As we'll see in chapter 10, survey questions are pretested before going into a questionnaire. If people won't give you straight answers to straight questions, you can back off and try an ordinal scale. But why start out crippling a perfectly good interval-scale question by making it ordinal when you don't know that you have to?

During data analysis you can lump interval-level data together into ordinal or nominal categories. If you know the ages of your respondents on a survey, you can divide them into "old" and "young"; if you know the number of calories consumed per week for each family in a study, you can divide the data into low, medium, and high. But you cannot do this trick the other way around. If you collect data on income by asking people whether they earn "up to a million pesos per year" or "more than a million per year," you cannot go back and assign actual numbers of pesos to each informant.

Notice that "up to a million" and "more than a million" is an ordinal variable that *looks like* a nominal variable because there are only two attributes. If the attributes are rankable, then the variable is ordinal. "A lot of fish" is more than "a small amount of fish," and "highly educated" is greater than "poorly educated." Ordinal variables can have any number of ranks. For purposes of statistical analysis, though, ordinal scales with five or more ranks are often treated as if they were interval-level variables. More about this in chapter 20 when we get to data analysis.

Units of Analysis

One of the very first things to do in any research project is decide on the **unit of analysis**. In a case study, there is exactly one unit of analysis—the village, the school, the hospital, the organization. Research designed to test

hypotheses requires many units of analysis, usually a sample from a large population—Navajos, Chicano migrants, Yanomami warriors, women in trade unions in Rio de Janeiro, runaway children who live on the street, people who go to chiropractors, Hispanic patrol officers in the U.S. Immigration and Naturalization Service who work on the border between the United States and Mexico.

Although most research in social science is about populations of people, many other things can be the units of analysis. You can focus on farms instead of farmers, or on unions instead of union members, or on wars instead of warriors. You can study marriage contracts; folk tales, songs, and myths; and countries, cultures, and cities.

Paul Doughty (1979), for example, surveyed demographic data on 134 countries in order to make a list of “primate cities.” Geographers say that a country has a primate city if its most populous city is at least twice the size of its second-most populous city. Doughty, an anthropologist who had worked in Peru, looked at the population of the three largest cities in each country and coded whether the largest city was at least three times greater than the second and third cities combined. He discovered that this extreme form of population concentration was associated with Latin America more than with any other region of the world at the time.

Holly Mathews (1985) did a study of how men and women in a Mexican village tell a famous folktale differently. The tale is called *La Llorona* (The Weeping Woman) and is known all over Mexico. Mathews’s research has to do with the problem of intracultural variation—different people telling the same story in different ways. She studied a sample of the population of *La Llorona* stories in a community where she was working. Each story, as told by a different person, had characteristics that could be compared across the sample of stories. One of the characteristics was whether the story was told by a man or by a woman, and this turned out to be the most important variable associated with the stories, which were the units of analysis. (See the section on schema analysis in chapter 17 for more about Mathews’s study of the *La Llorona* tales.)

You can have more than one unit of analysis in a study. When Mathews looked for similarities and differences in tellings of the story, then the stories were the units of analysis. But when she looked at patterns in the tellers of the stories, then people were her units of analysis.

Robert Augner (2004:145–162) asked 424 people in four ethnic groups (Sudanic, Efe, Bantu, and Tswa) in the Ituri Forest (Democratic Republic of Congo) about food taboos. For each of 145 animals, Augner asked each informant if it was edible, and if so, if there were any times when it should not be eaten. For example, some animals were said to be off limits to pregnant

women or to children; some animals required permission from an elder to eat; some animals should not be eaten by members of this or that clan; and so on. Aunger has data, then on 145 animals. When he analyzes those data and looks at which animals have similar patterns of avoidance, the animals are the units of analysis. But he also knows something about each of his 424 informants. When he looks at differences in food taboos across people—like patterns of food taboos in the four ethnic groups—then people are the units of analysis.

A Rule about Units of Analysis

Remember this rule: No matter what you are studying, always collect data on the lowest level unit of analysis possible.

Collect data about individuals, for example, rather than about households. If you are interested in issues of production and consumption (things that make sense at the household level), you can always package your data about individuals into data about households during analysis. But if you want to examine the association between female income and child spacing and you collect income data on households in the first place, then you are locked out. You can always **aggregate** data collected on individuals, but you can never **disaggregate** data collected on groups.

This rule applies whether you're studying people or countries. If you are studying relations among trading blocs in major world regions, then collect trade data on countries and pairs of countries, not on regions of the world.

Sometimes, though, *the smallest unit of analysis is a collective*, like a household or a region. For example, each person in a household consumes a certain number of grams of protein per week. But you can't just add up what individuals consume and get the number of grams of protein that comes into a household. Some grams are lost to waste, some to pets, some to fertilizer, some to fuel. After you add up all the grams, you get a single number for the household. If you are testing whether this number predicts the number of days per year that people in the household are sick, then the household is your unit of analysis.

The Ecological Fallacy

Once you select your unit of analysis, remember it as you go through data analysis, or you're likely to commit the dreaded "ecological fallacy." This fallacy (also known as the Nosnibor effect, after Robinson [1950], who described it) comes from drawing conclusions about the wrong units of analysis—making generalizations about people, for example, from data about groups or places. For example, in 1930, 11% of foreign-born people in the

United States were illiterate, compared with 3% of those born in the United States. The correlation between these two variables appeared to be positive. In other words, across 97 million people (the population of the United States at the time), being foreign born was a moderately strong predictor of being illiterate. But when Robinson looked at the data for the (then) 48 states in the United States, he got an entirely different result. The correlation between the percent illiterate and the percent of foreign-born people was $-.526$. That minus sign means that the more foreign born, the *less* illiteracy.

What's going on? Well, as Jargowsky (2005) observes, immigrants went mostly to the big industrial states where they were more likely to find jobs. Those northern and midwestern states had better schools and, of course, higher literacy—along with a lot of immigrants, many of whom were illiterate. And that was Robinson's point: if you only looked at the state-by-state averages (the aggregated units of analysis) instead of at the individual data, you'd draw the wrong conclusion about the relationship between the two variables. (For reviews of the ecological inference problem, see King 1997, Freedman 2001, and Jargowsky 2005.)

This is an important issue for anthropologists. Suppose you do a survey of villages in a region of southern India. For each village, you have data on such things as the number of people, the average age of men and women, and the monetary value of a list of various consumer goods in each village. That is, when you went through each village, you noted how many refrigerators and kerosene lanterns and radios there were, but you do not have these data for each person or household in the village because you were not interested in that when you designed your study. (You were interested in characteristics of villages as units of analysis.)

In your analysis, you notice that the villages with the population having the lowest average age also have the highest average dollar value of modern consumer goods. You are tempted to conclude that young people are more interested in (and purchase) modern consumer goods more frequently than do older people.

But you might be wrong. Villages with greater employment resources (land and industry) will have lower levels of labor migration by young people. Because more young people stay there, this will lower the average age of wealthier villages. Though *everyone* wants household consumer goods, only older people can afford them, having had more time to accumulate the funds.

It might turn out that the wealthy villages with low average age simply have wealthier older people than villages with higher average age. It is not valid to take data gathered about villages and draw conclusions about villagers, and this brings us to the crucial issue of validity.

Validity, Reliability, Accuracy, and Precision

Validity refers to the accuracy and trustworthiness of instruments, data, and findings in research. Nothing in research is more important than validity.

The Validity of Instruments and Data

Are the instruments that were used to measure something valid? Are SAT and GRE scores, for example, valid instruments for measuring the ability of students to get good grades? If they are, then are grades a valid measure of how smart students are?

Is the question “Do you practice polytheistic fetishism?” a valid instrument for measuring religious practices? No, it isn’t, because the concept of “polytheistic fetishism” is something that is meaningful only to specialists in the comparative study of religion. Asking people that question is asking them to think in categories that are alien to their culture.

Is the instrument “How long does it take you to drive to work each day?” a valid one for measuring the amount of time it takes people to drive to work each day? Well, that depends on how accurate you want the data to be. If you want the data to be accurate to within, say, 20 minutes on, say 70% of occasions, then the instrument is probably valid. If you want the data to be accurate to, say, within 5 minutes on, say, 90% of occasions, then the instrument is probably not valid because people just can’t dredge up the information you want at that level of accuracy.

The validity of data is tied to the validity of instruments. If questions asking people to recall their behavior are not valid instruments for tapping into informants’ past behavior, then the data retrieved by those instruments are not valid, either.

The Validity of Findings

Assuming, however, that the instruments and data are valid, we can ask whether the findings and conclusions derived from the data are valid. Asian Americans generally get higher scores on the math part of the SATs (scholastic aptitude tests) than do other ethnic groups in the United States. Suppose that the SAT math test is a valid instrument for measuring the general math ability of 18 year olds in the United States. Is it valid to conclude that “Asians are better at math” than other people are? No, it isn’t. That conclusion can only be reached by invoking an unfounded, racist assumption about the influence of certain genes—particularly genes responsible for epicanthic eye folds—on the ability of people to do math.

Reliability

Reliability refers to whether or not you get the same answer by using an instrument to measure something more than once. If you insert a thermometer into boiling water at sea level, it should register 212 Fahrenheit each and every time. “Instruments” can be things like thermometers and scales, or they can be questions that you ask people.

Like all other kinds of instruments, some questions are more reliable for retrieving information than others. If you ask 10 people “Do the ancestors take revenge on people who don’t worship them?” don’t expect to get the same answer from everyone. “How many brothers and sisters do you have?” is a pretty reliable instrument (you almost always get the same response when you ask a person that question a second time as you get the first time), but “How much is your parents’ house worth?” is much less reliable. And “How old were you when you were toilet trained?” is just futile.

Precision

Precision is about the number of decimal points in a measurement. Suppose your bathroom scale works on an old-fashioned spring mechanism. When you stand on the scale, the spring is compressed. As the spring compresses, it moves a pointer to a number that signifies how much weight is being put on the scale. Let’s say that you really, truly weigh 156.625 pounds, to the nearest thousandth of a pound.

If you have an old analog bathroom scale like mine, there are five little marks between each pound reading; that is, the scale registers weight in fifths of a pound. In terms of precision, then, your scale is somewhat limited. The best it could possibly do would be to announce that you weigh “somewhere between 156.6 and 156.8 pounds, and closer to the former figure than to the latter.” In this case, you might not be too concerned about the error introduced by lack of precision.

Whether you care or not depends on the needs you have for the data. If you are concerned about losing weight, then you’re probably not going to worry too much about the fact that your scale is only precise to the nearest fifth of a pound. But if you’re measuring the weights of pharmaceuticals, and someone’s life depends on your getting the precise amounts into a compound, well, that’s another matter.

Accuracy

Finally, **accuracy**. Assume that you are satisfied with the level of precision of the scale. What if the spring were not calibrated correctly (there was an

error at the factory where the scale was built, or last week your overweight house guest bent the spring a little too much) and the scale were off? Now we have the following interesting situation: The data from this instrument are valid (it has already been determined that the scale is measuring weight—exactly what you think it’s measuring); they are reliable (you get the same answer every time you step on it); and they are precise enough for your purposes. But they are not *accurate*. What next?

You could see if the scale were always inaccurate in the same way. You could stand on it 10 times in a row, without eating or doing exercise in between. That way, you’d be measuring the same thing 10 different times with the same instrument. If the reading were always the same, then the instrument would at least be reliable, even though it wasn’t accurate. Suppose it turned out that your scale were always incorrectly lower by 5 pounds. This is called **systematic bias**. Then, a simple correction formula would be all you’d need in order to feel confident that the data from the instrument were pretty close to the truth. The formula would be:

$$\text{true weight} = \text{your scale weight} + 5 \text{ pounds.}$$

The scale might be off in more complicated ways, however. It might be that for every 10 pounds of weight put on the scale, an additional half-pound correction has to be made. Then the **recalibration** formula would be:

$$\text{true weight} = (\text{your scale weight}) + (\text{scale weight} / 10)(.5)$$

or

$$(\text{your scale weight}) \times (1.05)$$

That is, take the scale weight, divide by 10, multiply by half a pound, and add the result to the reading on your scale.

If an instrument is not precise enough for what you want to do with the data, then you simply have to build a more precise one. There is no way out. If it is precise enough for your research and reliable, but inaccurate in known ways, then a formula can be applied to correct for the inaccuracy.

The real problem is when instruments are inaccurate in unknown ways. The bad news is that this happens a lot. If you ask people how long it takes them to drive to work, they’ll tell you. If you ask people what they ate for breakfast, they’ll tell you that, too. Answers to both questions may be dead on target, or they may bear no useful resemblance to the truth. The good news is that informant accuracy is one of the methodological questions that social scientists have been investigating for years and on which real progress continues to be made (Bernard et al. 1984; Sudman et al. 1996; Schwarz 1999; Vadez et al. 2003).

Determining Validity

You may have noticed a few paragraphs back that I casually slipped in the statement that some scale had *already been determined* to be a valid instrument. How do we know that the scale is measuring weight? Maybe it's measuring something else. How can we be sure? Since we have to make concepts up to study them, there is no direct way to evaluate the validity of an instrument for measuring a concept. Ultimately, we are left to decide, on the basis of our best judgment, whether an instrument is valid or not.

We are helped in making that judgment by some tests for **face validity**, **content validity**, **construct validity**, and **criterion validity**.

Face Validity

Establishing **face validity** involves simply looking at the operational indicators of a concept and deciding whether or not, *on the face of it*, the indicators make sense. On the face of it, asking people “How old were you when you were toilet trained?” is not a valid way to get at this kind of information. A paper-and-pencil test about the rules of the road is not, on the face of it, a valid indicator of whether someone knows how to drive a car. But the paper-and-pencil test is probably a valid test for determining if an applicant for a driver's license can read road signs. These different instruments—the road test and the paper-and-pencil test—have face validity for measuring different things.

Boster (1985) studied how well the women of the Aguaruna Jívaro in Peru understood the differences among manioc plants. He planted some fields with different varieties of manioc and asked women to identify the varieties. This technique, or instrument, for measuring cultural competence has great face validity; most researchers would agree that being able to identify more varieties of manioc is a valid indicator of cultural competence in this domain.

Boster might have simply asked women to list as many varieties of manioc as they could. This instrument would not have been as valid, on the face of it, as having them identify actual plants that were growing in the field. There are just too many things that could interfere with a person's memory of manioc names, even if they were super competent about planting roots, harvesting them, cooking them, trading them, and so on.

Face validity is based on consensus among researchers: If everyone agrees that asking people “How old are you” is a valid instrument for measuring age, then, until proven otherwise, that question is a valid instrument for measuring age.

Content Validity

Content validity is achieved when an instrument has appropriate content for measuring a complex concept, or construct. If you walk out of a test and feel that it was unfair because it tapped too narrow a band of knowledge, your complaint is that the test lacked content validity.

Content validity is very, very tough to achieve, particularly for complex, multidimensional constructs. Consider, for example, what's involved in measuring a concept like strength of ethnic identity among, say, second-generation Mexican Americans. Any scale to assess this has to have components that deal with religion, language, socioeconomic status, sense of history, and gastronomy.

Religion: Mexican Americans tend to be mostly Roman Catholic, but a growing number of Mexicans are now Protestants. The migration of a few million of these converts to the United States over the next decade will have an impact on ethnic politics—and ethnic identity—within the Mexican American population.

Language: Some second-generation Mexican Americans speak almost no Spanish; others are completely bilingual. Some use Spanish only in the home; others use it with their friends and business associates.

Socioeconomic status: Many Mexican Americans are poor (about 36% of Hispanic households in the United States have incomes below \$25,000 a year), but many others are well off (about 15% have incomes above \$75,000 a year) (SAUS 2004–2005, table 683). People with radically different incomes tend to have different political and economic values.

Sense of history: Some so-called Mexican Americans have roots that go back to before the British Pilgrims landed at Plymouth Rock. The Hispanos (as they are known) of New Mexico were Spaniards who came north from the Spanish colony of Mexico. Their self-described ethnic identity is quite different from recent immigrants from Mexico.

Gastronomy: The last refuge of ethnicity is food. When language is gone (Spanish, Yiddish, Polish, Gaelic, Greek, Chinese . . .), and when ties to the “old country” are gone, burritos, bagels, pirogis, corned beef, mousaka, and lo-mein remain. For some second-generation Mexican Americans, cuisine is practically synonymous with identity; for others it's just part of a much larger complex of traits.

A valid measure of ethnic identity, then, has to get at all these areas. People's use of Spanish inside and outside the home and their preference for Mexican or Mexican American foods are good measures of *some* of the content of Mexican American ethnicity. But if these are the only questions you ask, then your measure of ethnicity has low content validity. (See Cabassa [2003] for an assessment of acculturation scales for Hispanics in the United States.)

“Life satisfaction” is another very complex variable, composed of several concepts—like “having sufficient income,” “a general feeling of well-being,” and “satisfaction with level of personal control over one’s life.” In fact, most of the really interesting things that social scientists study are complex constructs, things like “quality of life,” “socioeconomic class,” “ability of teenagers to resist peer pressure to smoke,” and so on.

Construct Validity

An instrument has high **construct validity** if there is a close fit between the construct it supposedly measures and actual observations made with the instrument. An instrument has high construct validity, in other words, if it allows you to infer that a unit of analysis (a person, a country, whatever) has a particular complex trait and if it supports predictions that are made from theory.

Scholars have offered various definitions of the construct of ethnicity, based on different theoretical perspectives. Does a particular measure of Mexican American ethnicity have construct validity? Does it somehow “get at,” or measure, the components of this complex idea?

Asking people “How old are you?” has so much face validity that you hardly need to ask whether the instrument gets at the construct of chronological age. Giving people an IQ test, by contrast, is controversial because there is so much disagreement about what the construct of intelligence is. In fact, lots of constructs in which we’re interested—intelligence, ethnicity, machismo, alienation, acculturation—are controversial and so are the measures for them. Getting people to agree that a particular *measure* has high construct validity requires that they agree that the construct is valid in the first place.

Criterion Validity: The Gold Standard

An instrument has high **criterion validity** if there is a close fit between the measures it produces and the measures produced by some other instrument that is known to be valid. This is the gold standard test.

A tape measure, for example, is known to be an excellent instrument for measuring height. If you knew that a man in the United States wore shirts with 35” sleeves, and pants with 34” cuffs, you could bet that he was over 6’ tall and be right more than 95% of the time. On the other hand, you might ask: “Why should I measure his cuff length and sleeve length in order to know *most of the time, in general*, how tall he is, when I could use a tape measure and know *all of the time, precisely* how tall he is?”

Indeed. If you want to measure someone's height, use a tape measure. Don't substitute a lot of fuzzy proxy variables for something that's directly measurable by known, valid indicators. But if you want to measure things like quality of life and socioeconomic class—things that don't have well-understood, valid indicators—then a complex measure will just have to do until something simpler comes along.

The preference in science for simpler explanations and measures over more complicated ones is called the principle of **parsimony**. It is also known as **Ockham's razor**, after William Ockham (1285–1349), a medieval philosopher who argued *Pluralitas non est ponenda sine necessitate*, or “don't make things more complicated than they need to be.”

You can tap the power of criterion validity for complex constructs with the **known group comparison** technique. If you develop a scale to measure political ideology, you might try it out on members of the American Civil Liberties Union and on members of the Christian Coalition of America. Members of the ACLU should get high “left” scores, and members of the CCA should get high “right” scores. If they don't, there's probably something wrong with the scale. In other words, the known-group scores are the criteria for the validity of your instrument.

A particularly strong form of criterion validity is **predictive validity**—whether an instrument lets you predict accurately something else you're interested in. “Stress” is a complex construct. It occurs when people interpret events as threatening to their lives. Some people interpret a bad grade on an exam as a threat to their whole life, while others just blow it off. Now, stress is widely thought to produce a lowered immune response and increase the chances of getting sick. A really good *measure* of stress, then, ought to predict the likelihood of getting sick.

Remember the life insurance problem? You want to predict whether someone is likely to die in the next 365 days in order to know how much to charge them in premiums. Age and sex tell you a lot. But if you know their weight, whether they smoke, whether they exercise regularly, what their blood pressure is, whether they have ever had any one of a list of diseases, and whether they test-fly experimental aircraft for a living, then you can predict—with a higher and higher degree of accuracy—whether they will die within the next 365 days. Each piece of data—each component of a construct you might call “lifestyle”—adds to your ability to predict something of interest.

The Bottom Line

The bottom line on all this is that while various forms of validity can be demonstrated, Truth, with a capital T, is never final. We are never dead sure of

anything in science. We try to get closer and closer to the truth by better and better measurement. All of science relies on concepts whose existence must ultimately be demonstrated by their effects. You can ram a car against a cement wall at 50 miles an hour and account for the amount of crumpling done to the radiator by referring to a concept called “force.” The greater the force, the more crumpled the radiator. You demonstrate the existence of intelligence by showing how it predicts school achievement or monetary success.

The Problem with Validity

If you suspect that there is something deeply, desperately wrong with all this, you’re right. The whole argument for the validity (indeed, the very existence) of something like intelligence is, frankly, circular: How do you know that intelligence exists? Because you see its effects in achievement. And how do you account for achievement? By saying that someone has achieved highly because they’re intelligent. How do you know machismo exists? Because men dominate women in some societies. And how do you account for dominant behavior, like wife beating? By saying that wife beaters are acting out their machismo.

In the hierarchy of construct reality, then, force ranks way up there (after all, it’s got several hundred years of theory and experimentation behind it), while things like intelligence and machismo are pretty weak by comparison. And yet, as I made clear in chapter 1, the social and behavioral sciences are roaring successes, on a par with the physical sciences in terms of the effects they have on our lives every day. This is possible because social scientists have refined and tested many useful concepts and measurements for those concepts.

Ultimately, the validity of any concept—force in physics, the self in psychology, modernization in sociology and political science, acculturation in anthropology—depends on two things: (1) the utility of the device that measures it; and (2) the collective judgment of the scientific community that a concept and its measure are valid. In the end, we are left to deal with the effects of our judgments, which is just as it should be. Valid measurement makes valid data, but validity itself depends on the collective opinion of researchers.

Cause and Effect

Cause and effect is among the most highly debated issues in the philosophy of knowledge. (See Hollis [1996] for a review.) We can never be absolutely

certain that variation in one thing causes variation in another. Still, if measurements of two variables are valid, you can be reasonably confident that one variable causes another if four conditions are met.

1. The two variables **covary**—that is, as scores for one variable increase or decrease, scores for the other variable increase or decrease as well.
2. The covariation between the two variables is not **spurious**.
3. There is a **logical time order** to the variables. The presumed causal variable must always precede the other in time.
4. A mechanism is available that explains *how* an independent variable causes a dependent variable. There must, in other words, be a **theory**.

Condition 1: Covariation

When two variables are related they are said to **covary**. Covariation is also called **correlation** or, simply, **association**.

Association is a necessary but insufficient condition for claiming a causal relation between two variables. Whatever else is needed to establish cause and effect, you can't claim that one thing causes another if they aren't related in the first place.

Here are a few interesting covariations:

1. Sexual freedom for women tends to increase with the amount that women contribute to subsistence (Schlegel and Barry 1986).
2. Ground-floor, corner apartments occupied by students at big universities have a much higher chance of being burglarized than other units in the same apartment bloc (Robinson and Robinson 1997).
3. When married men and women are both employed full-time, they spend the same amount of time in the various rooms of their house—except for the kitchen (Ahrentzen et al. 1989).

You might think that in order to establish cause, independent variables would have to be strongly related to the dependent variable. Not always. People all over the world make decisions about whether or not to use (or demand the use of) a condom as a part of sexual relations. These decisions are based on many factors, all of which may be weakly, but causally related to the ultimate decision. These factors include: the education level of one or both partners; the level of income of one or both partners; the availability and cost of condoms; the amount of time that partners have been together; the amount of previous sexual experience of one or both partners; whether either or both partners know anyone personally who has died of AIDS; and so on.

Each independent variable may contribute only a little to the outcome of

the dependent variable (the decision that is finally made), but the contribution may be quite direct and causal.

Condition 2: Lack of Spuriousness

Just as weak correlations can be causal, strong correlations can turn out not to be. When this happens, the original correlation is said to be **spurious**. There is a strong correlation between the number of firefighters at a fire and the amount of damage done: the more firefighters, the higher the insurance claim. You could easily conclude that firefighters cause fire damage.

We know better: Both the amount of damage and the number of firefighters is caused by the size of the blaze. We need to **control** for this third variable—the size of the blaze—to understand what’s really going on.

Domenick Dellino (1984) found an inverse relation between perceived quality of life and involvement with the tourism industry on the island of Exuma in the Bahamas. When he controlled for the size of the community (he studied several on the island), the original correlation disappeared. People in the more congested areas were more likely to score low on the perceived-quality-of-life index whether or not they were involved with tourism, while those in the small, outlying communities were more likely to score high on the index. People in the congested areas were also more likely to be involved in tourism-related activities, because that’s where the tourists go.

Emmanuel Mwangi (1986) found that illiterates in Malawi were much more likely than literates to brew beer for sale from part of their maize crop. The covariation vanished when he controlled for wealth, which causes both greater education (hence, literacy) and the purchase, rather than the brewing, of maize beer.

The list of spurious relations is endless, and it is not always easy to detect them for the frauds that they are. A higher percentage of men than women get lung cancer, but when you control for the length of time that people have smoked, the gender difference in lung cancer vanishes. Pretty consistently, young people accept new technologies more readily than older people, but in many societies, the relation between age and readiness to adopt innovations disappears when you control for level of education. Urban migrants from tribal groups often give up polygyny in Africa and Asia, but both migration *and* abandonment of polygyny are often caused by a third factor: lack of wealth.

Your only defense against spurious covariations is vigilance. No matter how obvious a covariation may appear, discuss it with disinterested colleagues—people who have no stake at all in telling you what you want to hear. Present your initial findings in class seminars at your university or where you work.

Beg people to find potentially spurious relations in your work. You'll thank them for it if they do.

Condition 3: Precedence, or Time Order

Besides a nonspurious association, something else is required to establish a cause-and-effect relation between two variables: a **logical time order**. Firefighters don't cause fires—they show up *after* the blaze starts. African Americans have higher blood pressure, on average, than Whites do, but high blood pressure does not cause people to be African American.

Unfortunately, things are not always clear-cut. Does adoption of new technologies cause wealth, or is it the other way around? Does urban migration cause dissatisfaction with rural life, or the reverse? Does consumer demand cause new products to appear, or vice versa? Does the growth in the number of lawsuits cause more people to study law so that they can cash in, or does overproduction of lawyers cause more lawsuits?

What about the increase in elective surgery in the United States? Does the increased supply of surgeons cause an increase in elective surgery, or does the demand for surgery create a surfeit of surgeons? Or are both caused by external variables, like an increase in discretionary income in the upper middle class, or the fact that insurance companies pay more and more of Americans' medical bills?

Figure 2.4 shows three kinds of time order between two variables. Read figure 2.4(a) as “*a* is **antecedent** to *b*.” Read figure 2.4(b) as “*a* and *b* are antecedent to *c*.” And read figure 2.4(c) as “*a* is antecedent to *b*, which is an **intervening variable** antecedent to *c*.” A lot of data analysis is about understanding and controlling for antecedent and intervening variables—about which much more in chapter 20.

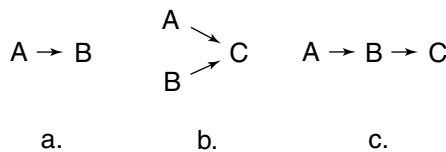


Figure 2.4. Time order between two or three variables.

Condition 4: Theory

Finally, even when you have established nonspurious, consistent, strong covariation, as well as a logical time sequence for two or more variables, you

need a **theory** that *explains* the association. Theories are good ideas about how things work.

One of my favorite good ideas about how things work is called **cognitive dissonance** theory (Festinger 1957). It's based on the insight that: (1) People can tell when their beliefs about what *ought* to be don't match their perception of how things really are; and (2) This causes an uncomfortable feeling. The feeling is called cognitive dissonance. People then have a choice: They can live with the dissonance (be uncomfortable); change the external reality (fight city hall); or change their beliefs (usually the path of least resistance, but not necessarily the easy way out).

Cognitive dissonance theory helps explain why some people accept new technologies that they initially reject out of fear for their jobs: Once a technology is entrenched, and there is no chance of getting rid of it, it's easier to change your ideas about what's good and what's bad than it is to live with dissonance (Bernard and Pelto 1987). Dissonance theory explains why some men change their beliefs about women working outside the home: When economic necessity drives women into the workforce, it's painful to hold onto the idea that that's the wrong thing for women to do.

On the other hand, some people do actually quit their jobs rather than accept new technologies, and some men continue to argue against women working outside the home, even when those men depend on their wives' income to make ends meet. This is an example of a general theory that fails to predict local phenomena. It leads us to seek more data and more understanding to predict when cognitive dissonance theory is insufficient as an explanation.

The literature is filled with good ideas for how to explain covariations. There is a well-known correlation between average daily temperature and the number of violent crimes reported to police (Anderson 1989; Cohn 1990). The association between temperature and violence, however, is neither as direct nor as simple as the correlational evidence might make it appear. Routine activity theory states that if you want to understand what people are doing, start with what they usually do. Social contact theory states that if you want to understand the probability for any event that involves human interaction, start by mapping activities that place people in contact with one another. Both of these theories are examples of Ockham's famous razor, discussed above.

Well, following routine activity theory, we find out that people are likely to be indoors, working, or going to school in air-conditioned comfort, during the hottest part of the day from Monday through Friday. Following social contact theory, we find that on very hot days, people are more likely to go out during the evening hours—which places them in more contact with one another. People also drink more alcohol during the evening hours. These facts, not temperature per se, may account for violence. Applying these theories, Cohn and

Rotton (1997) found that more crimes of violence *are* reported to police on hot days than on cool days, but those crimes are, in fact, more likely to occur during the cooler evening hours than during the hottest part of the day.

Many theories are developed to explain a purely local phenomenon and then turn out to have wider applicability. Many observers have noticed, for example, that when men from polygynous African societies move to cities, they often give up polygyny (Clignet 1970; Jacoby 1995). This consistent covariation is explained by the fact that men who move away from tribal territories in search of wage labor must abandon their land, their houses, and the shared labor of their kinsmen. Under those conditions, they simply cannot afford to provide for more than one wife, much less the children that multiple wives produce. The relation between urbanization and changes in marriage customs is explained by antecedent and intervening variables.

If you read the literature across the social sciences, you'll see references to something called "contagion theory." This one invokes a copycat mechanism to explain why suicides are more likely to come in batches when one of them is widely publicized in the press (Jamieson et al. 2003) and why more women candidates stand for election in districts that already have women legislators in office (Matland and Studlar 1996).

"Relative deprivation theory" is based on the insight that people compare themselves to specific peer groups, not to the world at large (Stouffer et al. 1949; Martin 1981). It explains why anthropology professors don't feel all that badly about engineering professors earning a lot of money, but hate it if sociologists in their university get significantly higher salaries. "World systems theory" proposes that the world's economies and political bodies are part of a single capitalist system that has a core and a periphery and that each nation can be understood in some sense by examining its place in that system (Wallerstein 1974, 2004).

All such theories start with one or two **primitive axioms**—things that are simply defined and that you have to take at face value. The definition of cognitive dissonance is an example: When people have inconsistent beliefs, or when they perceive things in the real world to be out of whack with their ideas of how things *should* be, they feel discomfort. This discomfort leads people to strive naturally toward cognitive consonance.

Neither the fact of dissonance, nor the discomfort it produces, nor the need for consonance are ever explained. They are primitive axioms. *How* people deal with dissonance and *how* they try to achieve consonance are areas for empirical research. As empirical research accumulates, the theory is tested and refined. William Dressler, a medical anthropologist, developed his theory of cultural consonance based on cognitive dissonance theory. Cultural consonance is the degree to which people's lives mirror a widely shared set of

beliefs about what lives should look like. What's a successful life? This differs from culture to culture, but in many cultures, the list of things that indicate success is widely shared. Dressler and his colleagues have found that people who have more of these things (whose lives are in consonance with the cultural model) have lower stress and fewer blood pressure problems than do people whose lives lack cultural consonance (Dressler et al. 1997, 2002; Dressler, Ribeiro et al. 2004, and see chapter 8 on measuring cultural consensus).

In relative deprivation theory, the fact that people have reference groups to which they compare themselves doesn't get explained, either. It, too, is a primitive axiom, an assumption, from which you deduce some results. The results are predictions, or hypotheses, that you then go out and test. The ideal in science is to deduce a prediction from theory and to test the prediction. That's the culture of science. The way social science really works much of the time is that you don't predict results, you **postdict** them. You analyze your data, come up with findings, and explain the findings after the fact.

There is nothing wrong with this. Knowledge and understanding can come from good ideas before you collect data or after you collect data. You must admit, though, there's a certain panache in making a prediction, sealing it in an envelope, and testing it. Later, when you take the prediction out of the envelope and it matches your empirical findings, you get a lot of points.

The Kalymnian Case

Here's an example of explaining findings after the fact. In my experience, it's pretty typical of how social scientists develop, refine, and change their minds about theories.

In my fieldwork in 1964–1965 on the island of Kalymnos, Greece, I noticed that young sponge divers (in their 20s) were more likely to get the bends than were older divers (those over 30). (The bends is a crippling malady that affects divers who come up too quickly after a long time in deep water.) I also noticed that younger divers were more productive than very old divers (those over 45), but not more productive than those in their middle years (30–40).

As it turned out, younger divers were subject to much greater social stress to demonstrate their daring and to take risks with their lives—risks that men over 30 had already put behind them. The younger divers worked longer under water (gathering more sponges), but they came up faster and were consequently at higher risk of bends. The middle group of divers made up in experience for the shortened time they spent in the water, so they maintained their high productivity at lower risk of bends. The older divers were feeling the

effects of infirmity brought on by years of deep diving, hence their productivity was lowered, along with their risk of death or injury from bends.

The real question was: What *caused* the young Kalymnian divers to engage in acts that placed them at greater risk?

My first attempt at explaining all this was pretty lame. I noticed that the men who took the most chances with their lives had a certain rhetoric and swagger. They were called *levédhis* (Greek for a brave young man) by other divers and by their captains. I concluded that somehow these men had more *levedhiá* (the quality of being brave and young) and that this made them higher risk takers. In fact, this is what many of my informants told me. Young men, they said, feel the need to show their manhood, and that's why they take risks by staying down too long and coming up too fast.

The problem with this cultural explanation was that it just didn't explain anything. Yes, the high risk takers swaggered and exhibited something we could label machismo or *levedhiá*. But what good did it do to say that lots of machismo caused people to dive deep and come up quickly? Where did young men get this feeling, I asked? "That's just how young men are," my informants told me. I reckoned that there might be something to this testosterone-poisoning theory, but it didn't seem adequate.

Eventually, I saw that the swaggering behavior and the values voiced about manliness were cultural ways to ratify, not explain, the high-risk diving behavior. Both the diving behavior and the ratifying behavior were the product of a third factor, an economic distribution system called *plátika*.

Divers traditionally took their entire season's expected earnings in advance, before shipping out in April for the 6-month sponge fishing expedition to North Africa. By taking their money (*plátika*) in advance, they placed themselves in debt to the boat captains. Just before they shipped out, the divers would pay off the debts that their families had accumulated during the preceding year. By the time they went to sea, the divers were nearly broke and their families started going into debt again for food and other necessities.

In the late 1950s, synthetic sponges began to take over the world markets, and young men on Kalymnos left for overseas jobs rather than go into sponge fishing. As divers left the island, the remaining divers demanded higher and higher *plátika*. They said that it was to compensate them for increases in the cost of living, but their demand for more money was a pure response by the divers to the increasing scarcity of their labor. The price of sponges, however, was dropping over the long term, due to competition with synthetics, so the higher *plátika* for the divers meant that the boat captains were losing profits. The captains put more and more pressure on the divers to produce more sponges, to stay down longer, and to take greater risks. This resulted in more accidents on the job (Bernard 1967, 1987).

Note that in all the examples of theory I've just given, the predictions and the post hoc explanations, I didn't have to quote a single statistic—not even a percentage score. That's because theories are qualitative. Ideas about cause and effect are based on insight; they are derived from either qualitative or quantitative observations and are initially expressed in words. *Testing* causal statements—finding out *how much* they explain rather than *whether* they seem to be plausible explanations—requires quantitative observations. But theory construction—explanation itself—is the quintessential qualitative act.

3



Preparing for Research

Setting Things Up

This chapter and the next are about some of the things that go on *before* data are collected and analyzed. I'll take you through the ideal research process and compare that to how research really gets done. Then I'll discuss the problem of choosing problems—how do I know what to study? In the next chapter, I'll give you some pointers on how to scour the literature so you can benefit from the work of others when you start a research project.

I'll have a lot more to say about the ethics of social research in this chapter—choosing a research problem involves decisions that can have serious ethical consequences—and a lot more about theory, too. Method and theory, it turns out, are closely related.

The Ideal Research Process

Despite all the myths about how research is done, it's actually a messy process that's cleaned up in the reporting of results. Figure 3.1 shows how the research process is supposed to work in the ideal world:

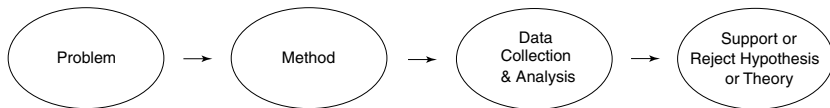


Figure 3.1. How research is supposed to work.

1. First, a theoretical problem is formulated;
2. Next, an appropriate site and method are selected;
3. Then, data are collected and analyzed;
4. Finally, the theoretical proposition with which the research was launched is either challenged or supported.

In fact, all kinds of practical and intellectual issues get in the way of this neat scheme. In the end, research papers are written so that the chaotic aspects of research are not emphasized and the orderly inputs and outcomes are.

I see nothing wrong with this. It would be a monumental waste of precious space in books and journals to describe the *real* research process for every project that's reported. Besides, every seasoned researcher knows just how messy it all is, anyway. You shouldn't have to become a highly experienced researcher before you're let into the secret of how it's really done.

A Realistic Approach

There are five questions to ask yourself about every research question you are thinking about pursuing. Most of these can also be asked about potential research sites and research methods. If you answer these questions honestly (at least to yourself), chances are you'll do good research every time. If you cheat on this test, even a teeny bit, chances are you'll regret it. Here are the five questions:

1. Does this topic (or research site, or data collection method) really interest me?
2. Is this a problem that is amenable to scientific inquiry?
3. Are adequate resources available to investigate this topic? To study this population at this particular research site? To use this particular data collection method?
4. Will my research question, or the methods I want to use, lead to unresolvable ethical problems?
5. Is the topic of theoretical and/or practical interest?

Personal Interest

The first thing to ask about any research question is: Am I really excited about this? Researchers do their best work when they are genuinely having fun, so don't do boring research when you can choose any topic you like.

Of course, you can't always choose any topic you like. In contract research, you sometimes have to take on a research question that a client finds interesting but that you find deadly dull. The most boring research I've ever done was on a contract where my coworkers and I combined ethnographic and survey

research of rural homeowners' knowledge of fire prevention and their attitudes toward volunteer fire departments. This was in 1973. I had young children at home and the research contract paid me a summer salary. It was honest work and I delivered a solid product to the agency that supported the project. But I never wrote up the results for publication.

By comparison, that same year I did some contract research on the effects of coed prisons on homosexuality among male and female inmates. I was very interested in that study and it was much easier to spend the extra time and effort polishing the contract reports for publication (Killworth and Bernard 1974).

I've seen many students doing research for term projects, M.A. theses, and even doctoral dissertations simply out of convenience and with no enthusiasm for the topic. If you are not interested in a research question, then no matter how important other people tell you it is, don't bother with it. If others are so sure that it's a dynamite topic of great theoretical significance, let *them* study it.

The same goes for people and places. Agricultural credit unions and brokerage houses are both complex organizations. But they are very different kinds of places to spend time in, so if you are going to study a complex organization, check your gut first and make sure you're excited about where you're going. It's really hard to conduct penetrating, in-depth interviews over a period of several weeks to a year if you aren't interested in the lives of the people you're studying.

You don't need any justification for your interest in studying a particular group of people or a particular topic. Personal interest is . . . well, personal. So ask yourself: Will my interest be sustained there? If the answer is "no," then reconsider. Accessibility of a research site or the availability of funds for the conduct of a survey are pluses, but by themselves they're not enough to make good research happen.

Science vs. Nonscience

The next question is: Is this a topic that can be studied by the methods of science? If the answer is "no," then no matter how much fun it is, and no matter how important it seems, don't even try to make a scientific study of it. Either let someone else do it, or use a different approach.

Consider this empirical question: How often do derogatory references to women occur in the Old Testament? If you can come up with a good, operational definition of "derogatory," then you can answer this question by looking through the corpus of data and counting the instances that turn up. Pretty straightforward, descriptive science.

But consider this question: Does the Old Testament offer support for unequal pay for women today? This is simply not answerable by the scientific method. It is no more answerable than the question: Is Rachmaninoff's music better than Tchaikovsky's? Or: Should the remaining hunting-and-gathering bands of the world be preserved just the way they are and kept from being spoiled by modern civilization? Whether or not a study is a scientific one depends first on the nature of the question being asked, and *then* on the methods used.

I can't stress too often or too strongly that when I talk about using the scientific method I'm *not* talking about numbers. In science, whenever a research problem can be investigated with quantitative measurement, numbers are more than just desirable; they're required. On the other hand, there are many intellectual problems for which quantitative measures are not yet available. Those problems require qualitative measurement.

Descriptions of processes (skinning a goat, building a fireworks tower, putting on makeup, setting the table for Thanksgiving), or of events (funerals, Little League games, parades), or of systems of nomenclature (kinship terms, disease terms, ways to avoid getting AIDS) require words, not numbers. Dorothy Holland and Debra Skinner (1987) asked some university women to list the kinds of guys there are. They got a list of words like "creep," "hunk," "nerd," "jerk," "sweetie pie," and so on. Then they asked some women, for each kind: "Is this someone you'd like to date?" The yes-no answers are nominal—that is, qualitative—measurement.

We'll get back to this kind of systematic, qualitative data collection in chapter 11.

Resources

The next question to ask is whether adequate resources are available for you to conduct your study. There are three major kinds of resources: time, money, and people. What may be adequate for some projects may be inadequate for others. Be totally honest with yourself about this issue.

TIME

Some research projects take a few weeks or months, while others take years. It takes a year or more to do an ethnographic study of a culture that is very different from your own, but a lot of focused ethnography can be done much more quickly. Gwendolyn Dordick (1996) spent 3 months studying a homeless shelter for 700 men in New York City. She visited the shelter four times a week for 3 hours or more each time, and spent 4 days at the shelter from

morning until lights-out at 10 P.M. This was enough time for her to understand a great deal about life in the shelter, including how a group of just 15 men had coalesced into a ruling elite and how some men had formed faux marriages (that could, but did not necessarily, involve sex) to protect themselves and their few possessions from violence and thievery.

Much of today's applied anthropological research is done in weeks or months, using rapid assessment methods. Rapid assessment methods are the same ones that everyone else uses but they are done quickly, and we'll cover these methods in chapter 13. If you are doing research for a term project, the topic has to be something you can look at in a matter of a few months—and squeezing the research into a schedule of other classes, at that. It makes no sense to choose a topic that demands two semesters' work when you have one semester in which to do the research. The effort to cram 10 gallons of water into a 5-gallon can is futile and quite common. Don't do it.

MONEY

Many things come under the umbrella of money. Equipment is essentially a money issue, as is salary or subsistence for you and other people involved in the research. Funds for assistants, computer time, supplies, and travel have to be calculated before you can actually conduct a major research project. No matter how interesting it is to you, and no matter how important it may seem theoretically, if you haven't got the resources to use the right methods, skip it for now.

Naturally, most people do not have the money it takes to mount a major research effort. That's why there are granting agencies. Writing proposals is a special craft. It pays to learn it early. Research grants for M.A. research are typically between \$1,000 and \$5,000. Grants for doctoral research are typically between \$5,000 and \$25,000. If you spend 100 hours working on a grant proposal that brings you \$10,000 to do your research, that's \$100/hr for your time. If you get turned down and spend another 100 hours rewriting the proposal, that's still \$50 an hour for your time if you're successful. Pretty good pay for interesting work.

If your research requires comparison of two groups over a period of 12 months, and you only have money for 6 months of research, can you accomplish your research goal by studying one group? Can you accomplish it by studying two groups for 3 months each? Ask yourself whether it's worthwhile pursuing your research if it has to be scaled down to fit available resources. If the answer is "no," then consider other topics.

Does the research require access to a particular village? Can you gain access to that village? Will the research require that you interview elite mem-

bers of the society you are studying—like village elders, shamans, medical malpractice lawyers, Lutheran priests? Will you be able to gain their cooperation? Or will they tell you to get lost or, even worse, lead you on with a lot of clichés about their culture? It's better not to do the study in the first place than to wind up with useless data.

PEOPLE

“People” includes you and others involved in the research, as well as those whom you are studying. Does the research require that you speak Papiamentu? If so, are you willing to put in the time and effort to learn that language? Can the research be done effectively with interpreters? If so, are such people available at a cost that you can handle? Does the research require that you personally do multiple regression? If it does, are you prepared to acquire that skill?

Ethics

I wish I could give you a list of criteria against which you could measure the “ethicalness” of every research idea you ever come up with. Unfortunately, it's not so simple. What's popularly ethical today may become popularly unethical tomorrow, and vice versa. (This does *not* mean that all ethics are relative. But more on that later.) During World War II, lots of anthropologists worked for what would today be called the Department of Defense, and they were applauded as patriots for lending their expertise to the war effort. In the 1960s, anthropologists took part in Project Camelot, a project by the U.S. Army to study counterinsurgency in Latin America (Horowitz 1965). This caused a huge outpouring of criticism, and the American Anthropological Association produced its first statement on ethics—not a formal code, but a statement—in 1967, rejecting quite specifically the use of the word “anthropology” as a disguise for spying (Fluehr-Lobban 1998:175).

During the Vietnam War, anthropologists who did clandestine work for the Department of Defense were vilified by their colleagues, and in 1971 the AAA promulgated a formal code of ethics, titled *Principles of Professional Responsibility*. That document specifically forbade anthropologists from doing any secret research and asserted the AAA's right to investigate allegations of behavior by anthropologists that hurts people who are studied, students, or colleagues (ibid.:177; see Wakin [1992] for details on anthropologists' work on counterinsurgency in Thailand during the Vietnam War).

Despite the rhetoric, though, no anthropologists have been expelled from

the AAA because of unethical conduct. One reason is that, when push comes to shove, everyone recognizes that there are conflicting, legitimate interests. In applied anthropology, for example, you have a serious obligation to those who pay for research. This obligation may conflict with your obligation to those whom you are studying. And when this happens, where do you stand? The Society for Applied Anthropology has maintained that the first obligation is to those whom we study. But the National Association of Practicing Anthropologists has promulgated a statement of professional responsibilities that recognizes how complex this issue can be.

We are a long, long way from finding the answers to these questions (Fluehr-Lobban 2002; Caplan 2003). Today, anthropologists are once again working for the Department of Defense. Is this simply because that's where the jobs are?

Perhaps. Times and popular ethics change. Whether you are subject to those changes is a matter for your own conscience, but it's because popular ethics change that Stanley Milgram was able to conduct his famous experiment on obedience in 1963.

Milgram's Obedience Experiment

Milgram duped people into thinking that they were taking part in an experiment on how well human beings learn under conditions of punishment. The subjects in the experiment were "teachers." The "learners" were Milgram's accomplices.

The so-called learners sat behind a wall, where they could be heard by subjects, but not seen. The subjects sat at a panel of 30 switches. Each switch supposedly delivered 30 more volts than the last, and the switches were clearly labeled from "Slight Shock" (15 volts) all the way up to "Danger: Severe Shock" (450 volts). Each time the learner made a mistake on a word-recall test, the subject was told to give the learner a bigger shock.

Milgram paid each participant \$4.50 up front (about \$30 in 2006 dollars). That made them feel obligated to go through with the experiment in which they were about to participate. He also gave them a little test shock—45 volts (the second lever on the 30-lever panel). That made people believe that the punishment they'd be delivering to the so-called learners was for real.

At 75 volts, the learner just grunted, but the reaction escalated as the putative voltage increased. At 150 volts, learners began pleading to be let out of the experiment. And at 285 volts, the learner's response, as Milgram reported it, could "only be described as an agonizing scream" (1974:4).

All those reactions by the learners were played back from tape so that subjects would hear the same things. The experimenter, in a white lab coat, kept

telling the subject to administer the shocks—saying things like: “You have no choice. You must go on.” A third of the subjects obeyed orders and administered what they thought were lethal shocks. Many subjects protested, but were convinced by the researchers in white coats that it was all right to follow orders.

Until Milgram did his troubling experiments (he did many of them, under different conditions and in different cities), it had been very easy to scoff at Nazi war criminals, whose defense was that they were “just following orders.” Milgram’s experiment taught us that perhaps a third of Americans had it in them to follow orders until they killed innocent people.

Were Milgram’s experiments unethical? Did the people who participated in Milgram’s experiments suffer emotional harm when they thought about what they’d done? If you were among Milgram’s subjects who obeyed to the end, would you be haunted by this? This was one of the issues raised by critics at the time (see Murray 1980).

Of course, Milgram **debriefed** the participants. (That’s where you make sure that people who have just participated in an experiment know that it had all been make-believe, and you help them deal with their feelings about the experiment.) Milgram tested 369 people in his experiments (Milgram 1977a). A year after the experiments ended, he sent them each a copy of his report and a follow-up questionnaire. He got back 84% of the questionnaires: About 1% said they were sorry or very sorry to have taken part in the experiment; 15% said they were neutral about the whole thing; and 84% said that, after reading the report and thinking about their experience, they were glad or very glad to have taken part in the experiment (Milgram 1977b).

Thomas Murray, a strong critic of deception in experiments, dismisses the idea that debriefing is sufficient. He points out that most social psychologists get very little training on how actually to conduct a debriefing and help people through any emotional difficulties. “Debriefings,” he says, “are more often viewed as discharging a responsibility (often an unpleasant one), an opportunity to collect additional data, or even as a chance for further manipulation!” (Murray 1980:14; but see Herrera [2001] for a critique of the critics of Milgram).

I can’t imagine Milgram’s experiment getting by a Human Subjects Review Committee at any university in the United States today, given the current code of ethics of the American Psychological Association (see appendix F). Still, it was less costly, and more ethical, than the natural experiments carried out at My Lai, or Chatilla—the Vietnamese village (in 1968) and the Lebanese refugee camps (in 1982)—whose civilian inhabitants were wiped out by American and Lebanese soldiers, respectively, “under orders.” Those experiments, too,

showed what ordinary people are capable of doing—except that in those cases, real people really got killed.

What Does It All Mean?

Just because times, and ethics, seem to change, does not mean that anything goes. Everyone agrees that scholars have ethical responsibilities, but not everyone agrees on what those responsibilities are. All the major scholarly societies have published their own code of ethics—all variations on the same theme, but all variations nonetheless. I've listed the Internet addresses for several of these codes of ethics in appendix F.

These documents are not perfect, but they cover a lot of ground and are based on the accumulated experience of thousands of researchers who have grappled with ethical dilemmas over the past 50 years. Look at those codes of ethics regularly during the course of any research project, both to get some of the wisdom that has gone into them and to develop your own ideas about how the documents might be improved.

Don't get trapped into nihilistic relativism. Cultural relativism (the unsalvageable fact that people's ideas about what is good and beautiful are shaped by their culture) is a great antidote for overdeveloped ethnocentrism. But, as Merrilee Salmon makes clear (1997), ethical relativism (that all ethical systems are equally good since they are all cultural products) is something else entirely.

Can you imagine defending the human rights violations of Nazi Germany as just another expression of the richness of culture? Would you feel comfortable defending, on the basis of cultural relativism, the so-called ethnic cleansing in the 1990s of Bosnians and Kosovar Albanians by Serbs in the former Yugoslavia? Or the slaughter of Tutsis by Hutus in Rwanda? Or of American Indians by immigrant Europeans 120 years earlier?

There is no value-free science. Everything that interests you as a potential research focus comes fully equipped with risks to you and to the people you study. Should anthropologists do social marketing for a state lottery? Or is social marketing only for getting people to use condoms and to wash their hands before preparing food? Should anthropologists work on projects that raise worker productivity in developing nations if that means some workers will become redundant? In each case, all you can do (and *must* do) is assess the potential human costs and the potential benefits. And when I say "potential benefits," I mean not just to humanity in the abstract, but also to you personally.

Don't hide from the fact that you are interested in your own glory, your own career, your own advancement. It's a safe bet that your colleagues are

interested in their career advancement, too. We have all heard of cases in which a scientist put her or his career above the health and well-being of others. This is devastating to science, and to scientists, but it happens when otherwise good, ethical people (1) convince themselves that they are doing something noble for humanity, rather than for themselves and (2) consequently fool themselves into thinking that *that* justifies their hurting others. (See Bhattacharjee [2004] for more on fraud in science.) When you make these assessments of costs and benefits, be prepared to come to decisions that may not be shared by all your colleagues. Remember the problem of the relation between darkness of skin color and measures of life success, like wealth and longevity? Would you, personally, be willing to participate in a study of this problem?

Suppose the study was likely to show that a statistically significant percentage of the variation in earning power in the United States is predictable from (*not* caused by) darkness of skin color. Some would argue that this would be useful evidence in the fight against racism and would jump at the chance to do the investigation. Others would argue that the evidence would be used by racists to do further damage in our society, so the study should simply not be done lest the information it produces fall into the wrong hands.

There is no answer to this dilemma. Above all, be honest with yourself. Ask yourself: Is this ethical? If the answer is “no,” then skip it; find another topic. Once again, there are plenty of interesting research questions that won’t put you into a moral bind. (For work on ethical issues of particular interest to anthropologists, see Harrison 1997, Cantwell et al. 2000, Fluehr-Lobban 2002, MacClancy 2002, Caplan 2003, Posey 2004, and Borofsky 2005.)

Theory: Explanation and Prediction

All research is specific. Whether you conduct ethnographic or questionnaire research, the first thing you do is *describe a process* or *investigate a relation* among some variables in a population. Description is essential, but to get from description to theory is a big leap. It involves asking: “What causes the phenomenon to exist in the first place?” and “What does this phenomenon cause?” Theory, then, is about explaining and predicting things.

It may seem odd to talk about theory in a book about methods, but you can’t design research until you choose a research question, and research questions depend crucially on theory. A good way to understand what theory is about is to pick something that begs to be explained and to look at competing explanations for it. See which explanation you like best. Do that for a few phenomena and you’ll quickly discover which paradigm you identify with.

That will make it easier to pick research problems and to develop hypotheses that you can go off and test.

Here is an example of something that begs to be explained: Everywhere in the world, there is a very small chance that children will be killed or maimed by their parents. However, the chance that a child is killed by a parent is much higher if a child has one or more nonbiological parents than if the child has two biological parents (Lightcap et al. 1982; Daly and Wilson 1988). All those evil-stepparent folktales appear to be based on more than fantasy.

Alternative Paradigms for Building Theories

One explanation is that this is all biological—in the genes. After all, male gorillas are known to kill off the offspring of new females they bring into their harem. Humans, the reasoning goes, have a bit of that instinct in them, too. They mostly fight and overcome the impulse, but over millions of cases, it's bound to come out once in a while. Culture usually trumps biology, but sometimes, biology is just stronger. This is a **sociobiological** explanation.

Another explanation is that it's cultural. Yes, it's more common for children to be killed by nonbiological than by biological parents, but this kind of mayhem is more common in some cultures than in others. Furthermore, although killing children is rare everywhere, in some cultures mothers are more likely to kill their children, while in other cultures fathers are more likely to be the culprits. This is because women and men learn different gender roles in different societies. So, the reasoning goes, we have to look at cultural differences for a true explanation of the phenomenon. This is called an **idealist**, or a cultural, theory because it is based on what people think—on their ideas.

Yet another explanation is that when adult men and women bring children to a second marriage, they know that their assets are going to be diluted by the claims the spouse's children have on those assets—immediate claims and claims of inheritance. This leads some of those people to harm their spouse's children from the former marriage. In a few cases, this causes death. This is a **materialist** theory.

Sociobiology, idealism, and materialism are not theories. They are **paradigms** or **theoretical perspectives**. They contain a few basic *rules for finding theories* of events. Sociobiology stresses the primacy of evolutionary, biological features of humans as the basis for human behavior. Idealism stresses the importance of internal states—attitudes, preferences, ideas, beliefs, values—as the basis for human behavior. And materialism stresses structural and infrastructural forces—like the economy, the technology of production, demography, and environmental conditions—as causes of human behavior.

When you want to explain a specific phenomenon, you apply the principles of your favorite paradigm and come up with a specific explanation.

Why do women everywhere in the world tend to have nurturing roles? If you think that biology rules here, then you'll be inclined to support evolutionary theories about other phenomena as well. If you think economic and political forces cause values and behavior, then you'll be inclined to apply the materialist perspective in your search for explanations in general. If you think that culture—people's values—is of paramount importance, then you'll tend to apply the idealist perspective in order to come up with explanations. The different paradigms are not so much in competition as they are complementary, for different **levels of analysis**. The sociobiological explanation for the battering of nonbiological children is appealing for aggregate, evolutionary phenomena—the big, big picture. A sociobiological explanation addresses the question: What is the reproductive advantage of this behavior happening at all?

But we know that the behavior of hurting or killing stepchildren is not inevitable, so a sociobiological explanation can't explain why some step-parents hurt their children and others don't. A materialist explanation is more productive for addressing this question. Some stepparents who bring a lot of resources to a second marriage become personally frustrated by the possibility of having their wealth raided and diluted by their new spouse's children. The reaction would be strongest for stepparents who have competing obligations to support their biological children who are living with yet another family. These frustrations will cause *some* people to become violent, but not others.

But even this doesn't explain why a particular stepparent is supportive or unsupportive of his or her nonbiological children. At this level of analysis, we need a processual and psychological explanation, one that takes into account the particular historical facts of the case. Whatever paradigm they follow, all empirical anthropologists rely on ethnography to test their theories.

Handwerker (1996b), for example, found that stepparents in Barbados were, overall, no more likely to treat children violently than were biological parents. But the presence of a *stepfather* increased the likelihood that women battered their daughters and decreased the likelihood that women battered their sons. In homes with stepparents, women saw their daughters as potential competitors for resources available from their partner and they saw sons as potential sources of physical protection and income.

And there was more. Powerful women (those who had their own sources of income) protected their children from violence, treated them affectionately, and elicited affection for them from their man. The probability that a son experienced an affectionate relationship with a biological father rose with the length of time the two lived together, but only for sons who had powerful

mothers. Men battered powerless women and the children of powerless women, and powerless women battered their own children.

Is there a sociobiological basis for powerful spouses to batter powerless ones? Or is this all something that gets stimulated by material conditions, like poverty? More research is needed on this fascinating question, but I think the points here are clear: (1) Different paradigms produce different answers to the same question; and (2) A lot of really interesting questions may have intriguing answers that are generated from several paradigms.

Idiographic and Nomothetic Theory

Theory comes in two basic sizes: elemental, or **idiographic** theory and generalizing or **nomothetic** theory. An idiographic theory accounts for the facts in a single case. A nomothetic theory accounts for the facts in many cases. The more cases that a theory accounts for, the more nomothetic it is.

The distinction was first made by Wilhelm Windelband, a philosopher of science, in 1894. By the late 1800s, Wilhelm Dilthey's distinction between the **Naturwissenschaften** and **Geisteswissenschaften**—the sciences of nature and the sciences of the mind—had become quite popular. The problem with Dilthey's distinction, said Windelband, was that it couldn't accommodate the then brand-new science of psychology. The subject matter made psychology a Geisteswissenschaft, but the discipline relied on the experimental method, and this made it a Naturwissenschaft.

What to do? Yes, said Windelband, the search for reliable knowledge is, indeed, of two kinds: the sciences of law and the sciences of events, or, in a memorable turn of phrase, "the study of what always is and the study of what once was." Windelband coined the terms **idiographic** and **nomothetic** to replace Dilthey's Natur- and Geisteswissenschaften.

Organic evolution is governed by laws, Windelband observed, but the sequence of organisms on this planet is an event that is not likely to be repeated on any other planet. Languages are governed by laws, but any given language at any one time is an event in human linguistic life. The goal of the idiographic, or historical sciences, then, is to deliver "portraits of humans and human life with all the richness of their unique forms" (Windelband 1998 [1894]:16).

Windelband went further. Every causal explanation of an event—every idiographic analysis, in other words—requires some idea of how things happen at all. No matter how vague the idea, there must be nomothetic principles guiding idiographic analysis.

Windelband's formulation is a perfect description of what all natural scientists—vulcanologists, ornithologists, astronomers, ethnographers—do all the

time. They describe things; they develop deep understanding of the cases they study; and they produce explanations for individual cases based on nomothetic rules. The study of *a* volcanic eruption, of *a* species' nesting habits, of *a* star's death is no more likely to produce new nomothetic knowledge than is the study of *a* culture's adaptation to new circumstances. But the idiographic effort, based on the application of nomothetic rules, is required equally across all the sciences if induction is to be applied and greater nomothetic knowledge achieved.

Those efforts in psychology are well known: Sigmund Freud based his theory of psychosexual development on just a few cases. Jean Piaget did the same in developing his universal theory of cognitive development, as did B. F. Skinner in developing the theory of operant conditioning.

In anthropology, Lewis Henry Morgan (1877) and others made a brave, if ill-fated effort in the 19th century to create nomothetic theories about the evolution of culture from the study of cases at hand. The unilineal evolutionary theories they advanced were wrong, but the effort to produce nomothetic theory was *not* wrong. Franz Boas and his students made clear the importance of paying careful attention to the particulars of each culture, but Leslie White and Julian Steward did not reject the idea that cultures evolve. Instead, they advanced more nuanced theories about how the process works (see Steward 1949, 1955; White 1959).

And the effort goes on. Wittfogel (1957) developed his so-called hydraulic theory of cultural evolution—that complex civilizations, in Mexico, India, China, Egypt, and Mesopotamia developed out of the need to organize the distribution of water for irrigation—based on idiographic knowledge of a handful of cases. David Price (1995) studied a modern, bureaucratically organized water supply system in the Fayoum area of Egypt. The further downstream a farmer's plot is from an irrigation pump, the less water he is likely to get because farmers upstream divert more water than the system allows them legally to have. Price's in-depth, idiographic analysis of the Fayoum irrigation system lends support to Wittfogel's long-neglected theory because, says Price, it shows "how farmers try to optimize the disadvantaged position in which the state has placed them" (ibid.:107–108). Susan Lees (1986) showed how farmers in Israel, Kenya, and Sudan got around bureaucratic limitations on the water they were allotted. We need much more idiographic analysis, more explanations of cases, in order to test the limitations of Wittfogel's theory.

Julian Steward (1955) chose a handful of cases when he developed his theory of cultural evolution. Data from Tehuacán, Mexico, and Ali Kosh, Iran—six thousand miles and several thousand years apart—support Steward's nomothetic formulation about the multistage transition from hunting and

gathering to agriculture. (The sequences appear to be similar responses to the retreat of the last glacier of the Paleolithic.) As we get more comparisons, the big picture will either become more and more nomothetic or it will be challenged.

Idiographic Theory

As in all sciences, most theory in anthropology is idiographic. Here are three examples:

1. In 1977, the New Delhi police reported 311 deaths by kitchen fires of women, mostly young brides who were killed because their families had not delivered a promised dowry to the groom's family (Claiborne 1984). By 1987, the government of India reported 1,912 such "**dowry deaths**" of young women, and by 1997 the number was 6,975—over 19 per day (Dugger 2000). How to explain this phenomenon?

Daniel Gross (1992) theorized that the phenomenon is a consequence of female hypergamy (marrying up) and dowry. Families that can raise a large dowry in India can marry off their daughter to someone of greater means. This has created a bidding war, as the families of wealthier sons demand more and more for the privilege of marrying those sons.

Apparently, many families of daughters in India have gone into debt to accumulate the dowries. When they can't pay off the debt, some of the families of grooms have murdered the brides in faked "kitchen accidents," where kerosene stoves purportedly blow up. This gives the grooms' families a chance to get another bride whose families can deliver. (For more on this issue, see Van Willigen and Chana [1991] and Thakur [1996].)

2. Next, consider the well-known case of fraternal polyandry. Hiatt (1980) noticed that among the Sinhalese of Sri Lanka, there was a shortage of women among those groups that practiced polyandry. He theorized that the shortage of women accounted for the practice of polyandry.

Earlier, Goldstein (1971) had observed that in Tibet, polyandry was practiced only among people who didn't own land. It turns out that in feudal times, some peasants were given a fixed allotment of land which they could pass on to their sons. In order to not break up the land, brothers would take a single bride into one household.

3. Finally, consider an idiographic theory derived entirely from ethnography. Anthony Paredes has been doing research on the Poarch Band of Creek Indians in Alabama since 1971. When he began his research, the Indians were a remnant

of an earlier group. They had lost the use of the Creek language, were not recognized by the U.S. government as a tribe, and had little contact with other Indians for decades. Yet, the Poarch Creek Indians had somehow maintained their identity.

Paredes wanted to know how the Indians had managed this. He did what he called “old-fashioned ethnography,” including key-informant interviewing and learned about a cultural revitalization movement that had been going on since the 1940s. That movement was led by some key people whose efforts over the years had made a difference. Paredes’s description of how the Poarch Creek Indians held their cultural identity in the face of such odds is an excellent example of elemental, idiographic theory. As you read his account you feel you understand how it worked (see Paredes 1974, 1992).

SO WHAT’S WRONG?

Nothing’s wrong. Gross’s intuitively appealing explanation for the kitchen fires in India rings true, but it doesn’t explain why other societies that have escalating dowry don’t have kitchen fires. Nor does it tell us why dowry persists in India despite its being outlawed since 1961, or why dowry—which, after all, only occurs in 7.5% of the world’s societies—exists in the first place. But Gross’s theory is a first-class example of theory at the local level—where research begins.

Goldstein’s attractive theory explains the Tibetan case of fraternal polyandry, but it doesn’t explain other reported cases of polyandry, like the one Hiatt studied in Sri Lanka.

Paredes’s convincing theory of how the Poarch Creeks maintained their cultural identity doesn’t tell us how other Native American groups managed to do this or why some groups did *not* manage it. Nor does it tell us anything about why other ethnic groups maintain or fail to maintain their identity in the United States or why ethnicity persists at all in the face of pressure from states on ethnic groups to assimilate. Fine. Others can try to make the theory more nomothetic.

In any science, much of the best work is at the idiographic level of theory making.

Nomothetic Theory

Nomothetic theories address questions like “So, what *does* account for the existence of dowry?”

Several theorists have tried to answer this question. Esther Boserup (1970)

hypothesized that dowry should occur in societies where a woman's role in subsistence production is low. She was right, but many societies where women's productive effort is of low value do *not* have dowry.

Gaulin and Boster (1990) offered a sociobiological theory that predicts dowry in stratified societies that have monogamous or polyandrous marriage. They tested their theory on Murdock and White's (1969) Standard Cross-Cultural Sample of 186 societies. The Gaulin-Boster theory works better than Boserup's—it misclassifies fewer societies—but still makes some mistakes. Fully 77% of dowry societies are, in fact, stratified and have monogamous marriage, but 63% of all monogamous, stratified societies do *not* have dowry.

Marvin Harris (1980), building on Boserup's model, hypothesized that dowry should occur in societies where women's role in subsistence production is low *and* where their value in reproduction is also low. In other words, if women are a liability in both their productive and reproductive roles, one should expect dowry as a compensation to the groom's family for taking on the liability represented by a bride who marries into a groom's family.

Kenneth Adams (1993) operationalized this idea. He reasoned that, since women are less suited physically to handling a plow, societies with plow agriculture and high-quality agricultural land should find women's labor of low value. If those societies have high population density, then women's reproductive role should be of low value. Finally, in societies with both these characteristics, patrilocal residence would make accepting a bride a real liability and would lead to demand for compensation—hence, dowry.

Adams tested his idea on the same sample of 186 societies. His theory makes about 25% fewer errors than the Gaulin-Boster theory does in predicting which societies have dowry. There has thus been a succession of theories to account for dowry; each theory has done a bit better than the last; and each has been based on reasoning from common-sense principles. That's how nomothetic theory grows.

A lot of comparative research is about testing nomothetic theory. If an idiographic theory accounts for some data in say, India, Japan, or England, then an obvious next step is to see how far the theory extends. Alice Schlegel and Herbert Barry (1986), for example, looked at the consequences of female contribution to subsistence. Their nomothetic theory predicts that women will be more respected in societies where they contribute a lot to subsistence than in societies where their contribution is low.

Whether their theory is supported depends crucially on how Schlegel and Barry operationalize the concept of *respect*. In societies where women contribute a lot to subsistence, say Schlegel and Barry, women will be spared some of the burden of pregnancy “through the attempt to space children” more evenly (ibid.:146). In such societies, women will be subjected to rape

less often; they will have greater sexual freedom; they will be worth more in bride wealth; and they will have greater choice in selection of a spouse. Schlegel and Barry coded the 186 societies in the Standard Cross-Cultural Sample for each of those indicators of respect—and their predictions were supported.

ONE MORE: THE SECOND DEMOGRAPHIC TRANSITION

Let's do one more—the second demographic transition. The first demographic transition happened at the end of the Paleolithic when people swapped hunting and gathering for agriculture as the main means of production. During the Paleolithic, population growth was very, very slow. But across the world, as people switched from hunting and gathering to agriculture, as they settled down and accumulated surplus, their populations exploded.

The second demographic transition began in the late 18th century in Europe with industrialization and has been spreading around the world ever since. Today, Japan, Germany, Italy, and other highly industrialized countries have **total fertility rates** (the average number of children born to women during their reproductive years), or **TFRs**, in the neighborhood of 1.5 to 1.2—that's 29% to 43% below the 2.1 TFR needed in those countries just to replace the current population. In the last 30 years, some previously high TFR countries, like Barbados, Mauritius, and Mexico, have been through a major demographic transition.

Explaining why women in Barbados are having fewer children is idiographic; predicting the conditions under which women in *any* underdeveloped country will start lowering their fertility rate is nomothetic. Handwerker's theory (1989) is that women in low-wage jobs encourage their daughters to get more education. And when women get sufficiently educated, their participation in the labor market becomes more effective (they earn more), freeing them from dependency on men (sons and husbands). As this dependency diminishes, women lower their fertility.

Handwerker's theory is nomothetic and materialist. It relies on material conditions to explain how preferences develop for fewer children and it does not rely on preferences (culture, ideas, values) to explain the level of a country's TFR.

The Consequences of Paradigm

Differences in theoretical paradigms have profound consequences. If you think that beliefs and attitudes are what make people behave as they do, then if you want to change people's behavior, the obvious thing to do is change their attitudes. This is the basis of the **educational model of social change**

I mentioned in chapter 2—the runaway best-seller model for change in our society.

Want to get women in developing nations to have fewer children? Educate them about the importance of small families. Want to lower the rate of infectious disease in developing countries? Educate people about the importance of good hygiene. Want to get adolescents in Boston or Seattle or wherever to stop having high-risk sex? Educate them about the importance of abstinence or, if that fails, about how to take protective measures against sexually transmitted disease. Want to get people in the United States to use their cars less? Educate them about car pooling.

These kinds of programs rarely work—but they do work sometimes. You *can* educate people (through commercial advertising) about why they should switch from, say, a Honda to a Toyota, or from a Toyota to a Ford, but you can't get people to give up their cars. You *can* educate people (through social advertising) to use the pill rather than less-effective methods of birth control, once they have decided to lower their fertility, but educational rhetoric doesn't influence the number of children that people want in the first place.

The closer a behavior is to the culture (or **superstructure**) of society, the easier it is to intervene culturally. Brand preferences are often superstructural, so advertising works to get people to switch brands—to change their behavior. But if people's behavior is rooted in the **structure** or **infrastructure** of society, then forget about changing their behavior by educating them to have better attitudes.

Eri Sugita (2004) studied 50 women in rural Uganda who had children under 5 years of age in their care. Over 14 months of fieldwork, Sugita found that the number of liters of water available per person in each household did a better job of predicting whether the women washed their hands before preparing or serving food than did the women's education or knowledge of hygiene. Women and children fetched all the water, so those who lived near a well were able to make more trips and get more water—unless they had a bicycle. Men, however, monopolized the bicycle in families that could afford one. Teaching people more about hygiene wouldn't do nearly as much for public health in that village as giving women bicycles would.

In poor countries, having many children may be the only security people have in their old age. No amount of rhetoric about the advantages of small families is going to change anyone's mind about the number of children they want to have. If you need a car because the only affordable housing is 30 miles from your job, no amount of rhetoric will convince you to take the bus. Demographic transition theory is highly nomothetic. It accounts for why Japan, a fully industrialized nation, has such a low TFR. But it doesn't predict what the consequences of that low TFR will be. For the time being, at least (until

even bigger nomothetic theories are developed), we still need an idiographic theory for this.

Japan has about 130 million people—about half the population of the United States—living in an area the size of Montana. The Japanese enjoy one of the highest average per capita incomes in the world. This is based on manufacturing products for export. The oil to run the factories that produce all those exports has to be imported. So does a lot of food to feed all those people who are working in the factories. The TFR of 1.3 in Japan makes it easy to predict that, in the next 20 or 30 years, Japan's industries will need to find lots of new workers to maintain productivity—and the lifestyle supported by that productivity.

Belgium and Italy—two other countries with low TFRs—have solved this problem by opening their borders to people from the formerly communist countries of eastern Europe. There are lots of people in Asia who are looking for work, but 97% of Japan's population is ethnically Japanese. Many Japanese don't like the idea of opening their borders to, say, Filipinos or Indonesians, so it will be a hard sell, in terms of domestic politics, for the government of Japan that proposes this solution to the problem of the coming labor shortage. Japan could recruit women more fully into the workforce, but many Japanese—particularly men—find this unappealing as well.

Obviously, something will have to give. Either Japan's productivity will drop, or workers will be recruited from abroad, or Japanese women will be recruited into the high-paying jobs of Japan's industrial machine. Demographic transition theory—a very nomothetic theory, indeed—does not tell us which of these alternatives will win out. Idealist theorists, particularly scholars who are immersed in the realities of modern Japanese society and culture, will contribute their own ideas about which choice will win. What I'm hoping for is a nomothetic theory that explains this kind of choice in many countries, not just in one.

This is one of my current favorite topics to think about because it illustrates how important theory is in developing research questions and it showcases the contributions of idealist and materialist perspectives, as well as the importance of idiographic and nomothetic theory.

There is no “list” of research questions. You have to use your imagination and your curiosity about how things work, and follow your hunches. Above all, never take anything at face value. Every time you read an article, ask yourself: “What would a study look like that would test whether the major assertions and conclusions of this article were really correct?” If someone says: “The only things students really care about these days are sex, drugs, and rock-and-roll,” the proper response is: “We can test that.”

A Guide to Research Topics, Anyway

There may not be a list of research topics, but there are some useful guidelines. First of all, there are very few big-theory issues—I call them research *arenas*—in all of social science. Here are four of them: (1) the nature-nurture problem, (2) the evolution problem, (3) the internal-external problem, and (4) the social facts or emergent properties problem.

1. **The nature-nurture problem.** This is an age-old question: How much of our personality and behavior is determined by our genes and how much by our exposure to different environments? Many diseases (cystic fibrosis, Tay-Sachs, sickle-cell anemia) are completely determined by our genes, but others (heart disease, diabetes, asthma) are at least partly the result of our cultural and physical environment.

Schizophrenia is a genetically inherited disease, but its expression is heavily influenced by our cultural environment. Hallucinations are commonly associated with schizophrenia but when Robert Edgerton (1966) asked over 500 people in four East African tribes to list the behavior of people who are severely mentally ill, less than 1% of them mentioned hallucinations (see also Edgerton and Cohen 1994; Jenkins and Barrett 2004).

Research on the extent to which differences in cognitive functions of men and women are the consequence of environmental factors (nurture) or genetic factors (nature) or the interaction between those factors is part of this research arena (Caplan et al. 1997). So are studies of human response to signs of illness across cultures (Kleinman 1980; Hielscher and Sommerfeld 1985).

2. **The evolution problem.** Studies of how groups change through time from one *kind* of thing to another kind of thing are in this arena. Societies change very slowly through time, but at some point we say that a village has changed into a city or that a society has changed from a feudal to an industrial economy. All studies of the differences between small societies—**Gemeinschaften**—and big societies—**Gesellschaften**—are in this arena. So are studies of inexorable bureaucratization as organizations grow.
3. **The internal-external problem.** Studies of the way in which behavior is influenced by values and by environmental conditions are in this arena. Studies of **response effects** (how people respond differently to the same question asked by a woman or by a man, for example) are in this arena, too. So are studies of the difference between what people say they do and what they actually do.
4. **The social facts, or emergent properties problem.** The name for this problem comes from Emile Durkheim's (1933 [1893]) argument that social facts exist outside of individuals and are not reducible to psychological facts. A great deal of social research is based on the assumption that people are influenced by social

forces that *emerge* from the interaction of humans but that transcend individuals. Many studies of social networks and social support, for example, are in this arena, as are studies that test the influence of organizational forms on human thought and behavior.

Generating Types of Studies

Now look at table 3.1. I have divided research topics (not arenas) into classes, based on the relation among kinds of variables.

TABLE 3.1
Types of Studies

	<i>Internal States</i>	<i>External States</i>	<i>Reported Behavior</i>	<i>Observed Behavior</i>	<i>Artifacts</i>	<i>Environment</i>
Internal states	I	II	IIIa	IIIb	IV	V
External states		VI	VIIa	VIIb	VIII	IX
Reported behavior			Xa	Xb	XIa	XIIa
Observed behavior				Xc	XIb	XIIb
Artifacts					XIII	XIV
Environment						XV

The five major kinds of variables are:

1. **Internal states.** These include attitudes, beliefs, values, and perceptions. Cognition is an internal state.
2. **External states.** These include characteristics of people, such as age, wealth, health status, height, weight, gender, and so on.
3. **Behavior.** This covers what people eat, who they communicate with, how much they work and play—in short, everything that people do and much of what social scientists are interested in understanding.
4. **Artifacts.** This includes all the physical residue from human behavior—radioactive waste, tomato slices, sneakers, arrowheads, computer disks, Viagra, skyscrapers—everything.
5. **Environment.** This includes physical and social environmental characteristics. The amount of rainfall, the amount of biomass per square kilometer, location on a river or ocean front—these are physical features that influence human thought and behavior. Humans also live in a social environment. Living under a democratic vs. an authoritarian régime or working in an organization that tolerates or does not tolerate sexual harassment are examples of social environments that have consequences for what people think and how they behave.

Keep in mind that category (3) includes both reported behavior and actual behavior.

A great deal of research has shown that about a third to a half of everything

people report about their behavior is not true (Bernard et al. 1984). If you want to know what people eat, for example, asking them is not a good way to find out (Basiotis et al. 1987; Johnson et al. 1996). If you ask people how many times a year they go to church, you're likely to get highly exaggerated data (Hadaway et al. 1993, 1998).

Some of the difference between what people say they do and what they actually do is the result of out-and-out lying. Most of the difference, though, is the result of the fact that people can't hang on to the level of detail about their behavior that is called for when they are confronted by social scientists asking them how many times they did this or that in the last month. What people *think* about their behavior may be what you're interested in, but that's a different matter.

Most social research focuses on internal states and on reported behavior. But the study of humanity can be much richer, once you get the hang of putting together these five kinds of variables and conjuring up potential relations. Here are some examples of studies for each of the cells in table 3.1.

Cell I:

The interaction of internal states, like perceptions, attitudes, beliefs, values, and moods.

Religious beliefs and attitudes about conservation and the environment (Nooney et al. 2003).

Perceived gender role (as measured with the Bem Sex Role Inventory) and attitudes about rape in Turkey (Gölge et al. 2003).

This cell is also filled with studies that compare internal states across groups. For example, Cooke's (2004) study of attitudes toward gun control among American, British, and Australian youth.

Cell II:

The interaction of internal states (perceptions, beliefs, moods, etc.) and external states (completed education, health status, organizational conditions).

Health status and hopefulness about the future (Vieth et al. 1997).

The relation between racial attitudes and the political context in different cities (Glaser and Gilens 1997).

Cell IIIa:

The interaction between *reported* behavior and internal states.

Attitudes toward the environment and reported environment-friendly behavior (Minton and Rose 1997).

Reported rate of church attendance and attitude toward premarital sex (Petersen and Donnenwerth 1997).

Cell IIIb:

The interaction between *observed* behavior and internal states.

Attitudes and beliefs about resources and actual behavior in the control of a household thermostat (Kempton 1987).

The effect of increased overtime work on cognitive function in automotive workers, including attention and mood (Proctor et al. 1996).

Cell IV:

The interaction of material artifacts and internal states.

The effects on Holocaust Museum staff in Washington, D.C., of working with the physical reminders of the Holocaust (McCarroll et al. 1995).

The ideas and values that brides and grooms in the United States share (or don't share) about the kinds of ritual artifacts that are supposed to be used in a wedding (Lowrey and Otnes 1994).

How children learn that domestic artifacts are considered feminine while artifacts associated with nondomestic production are considered masculine (Crabb and Bielawski 1994).

Cell V:

The interaction of social and physical environmental factors and internal states.

How culture influences the course of schizophrenia (Edgerton and Cohen 1994).

The extent to which adopted children and biological children raised in the same household develop similar personalities (McGue et al. 1996).

Cell VI:

How the interaction among external states relates to outcomes, like longevity or financial success.

The effects of things like age, sex, race, marital status, education, income, employment status, and health status on the risk of dying from the abuse of illegal drugs (Kallan 1998).

The interaction of variables like marital status, ethnicity, medical risk, and level of prenatal care on low birth weight (Abel 1997).

The effect of skin color on acculturation among Mexican Americans (Vasquez et al. 1997).

Cell VIIa:

The relation between external states and *reported* behavior.

The likelihood that baby-boomers will report attending church as they get older (Miller and Nakamura 1996).

The effect of age, income, and season on how much leisure time Tawahka Indian spouses spend with each other (Godoy 2002).

Gender differences in self-reported suicidal behavior among adolescents (Vannatta 1996).

Cell VIIb:

The relation between external states and *observed* behavior.

Health status, family drug history, and other factors associated with women who successfully quit smoking (Jensen and Coombs 1994). (Note: This is also an example of Cell XIIIb.)

Observed recycling behavior among Mexican housewives is better predicted by their observed competencies than by their beliefs about recycling (Corral-Verdugo 1997).

Cell VIII:

The relation of physical artifacts and external states.
 How age and gender differences relate to cherished possessions among children and adolescents from 6 to 18 years of age (Dyl and Wapner 1996).
 How engineering drawings and machines delineate boundaries and facilitate interaction among engineers, technicians, and assemblers in a firm that manufactures computer chips (Bechky 2003).

Cell IX:

The relation of external states and environmental conditions.
 How the work environment contributes to heart disease (Kasl 1996).
 Relation of daily levels of various pollutants in the air and such things as violent crimes or psychiatric emergencies (Briere et al. 1983).
 How proximity to a supermarket affects the nutrition of pregnant women (Laraia et al. 2004).

Cell Xa:

The relation between behaviors, as *reported* by people to researchers.
 The relation of self-reported level of church attendance and self-reported level of environmental activism among African Americans in Louisiana (Arp and Boeckelman 1997).
 The relation of reported changes in fertility practices to reported changes in actions to avoid HIV infection among women in rural Zimbabwe (Gregson et al. 1998).

Cell Xb:

The relation between reported and *observed* behavior.
 Assessing the accuracy of reports by Tsimane Indians in Bolivia about the size of forest plots they've cleared in the past year by comparing those reports to a direct physical measure of the plots (Vadez et al. 2003).
 The relation of reports about recycling behavior and actual recycling behavior (Corral-Verdugo 1997).
 Comparing responses by frail elderly men to the Activities of Daily Living Scale with observations of those same men as they engage in activities of daily living (Skruppy 1993).

Cell XIa:

The relation of *observed* behavior to specific physical artifacts.
 Content analysis of top-grossing films over 31 years shows that "tobacco events" (which include the presence of tobacco paraphernalia, as well as characters talking about smoking or actually smoking) are disproportionate to the actual rate of smoking in the population (Hazan et al. 1994).

Cell XIb:

The relation of *reported* behavior to specific physical artifacts.

People who are employed view prized possessions as symbols of their own personal history, whereas people who are unemployed see prized possessions as having utilitarian value (Ditmar 1991).

Cell XIIa:

The relation of *reported* behavior to factors in the social or physical environment.

The relation of compulsive consumer behavior in young adults to whether they were raised in intact or disrupted families (Rindfleisch et al. 1997).

Cell XIIb:

The relation of *observed* behavior to factors in the social or physical environment.

People are willing to wait longer when music is playing than when there is silence (North and Hargreaves 1999).

How environmental features of gay bathhouses facilitate sexual activity (Tewksbury 2002).

Cell XIII:

The association of physical artifacts to one another and what this predicts about human thought or behavior.

The research on how to arrange products in stores to maximize sales is in this cell.

Comparing the favorite possessions of urban Indians (in India) and Indian immigrants to the United States to see whether certain sets of possessions remain meaningful among immigrants (Mehta and Belk 1991). This is also an example of Cell IV. Note the difference between expressed *preferences* across artifacts and the coexistence of artifacts across places or times.

Cell XIV:

The probability that certain artifacts (relating, for example, to subsistence) will be found in certain physical or social environments (rain forests, deserts, shoreline communities). This area of research is mostly the province of archeology.

Cell XV:

How features of the social and physical environment interact and affect human behavioral and cognitive outcomes.

Social and physical environmental features of retail stores interact to affect the buying behavior of consumers (Baker et al. 1992).

The above list is only meant to give you an idea of how to think about potential covariations and, consequently, about potential research topics. Always keep in mind that *covariation does not mean cause*. Covariation can be the result of an antecedent or an intervening variable, or even just an acci-

dent. (Refer to chapter 2 for a discussion of causality, spurious relations, and antecedent variables.)

And keep in mind that many of the examples in the list above are statements about possible **bivariate correlations**—that is, about possible covariation between two things. Social phenomena being the complex sorts of things they are, a lot of research involves **multivariate relations**—that is, covariation among three or more things at the same time.

For example, it's well known that people who call themselves religious conservatives in the United States are likely to support the National Rifle Association's policy on gun control (Cell I). But the association between the two variables (religious beliefs and attitudes toward gun control) is by no means perfect and is affected by many intervening variables.

I'll tell you about testing for bivariate relations in chapter 20 and about testing for multivariate relations in chapter 21. As in so many other things, you crawl before you run and you run before you fly.

4



The Literature Search

The first thing to do after you get an idea for a piece of research is to find out what has already been done on it. Don't neglect this very important part of the research process. You need to make a heroic effort to uncover sources. People will know it if you don't, and without that effort you risk wasting a lot of time going over already-covered ground. Even worse, you risk having your colleagues ignore your work because you didn't do your homework.

Fortunately, with all the new **documentation resources** available, heroic efforts are actually pretty easy. There are three main documentation resources: (1) people, (2) review articles and bibliographies, and (3) a host of online databases.

People

There is nothing useful, prestigious, or exciting about discovering literature on your own. Reading it is what's important, so don't waste any time in finding it. Experts are great documentation resources. Begin by asking everyone and anyone whom you think has a remote chance of knowing something about the topic you're interested in if they can recommend some key articles or books that will get you into the literature on your topic.

Use the network method to conduct this first stage of your literature review. If the people you know are not experts in the topic you're studying, ask them if they know personally any people who *are* experts. Then contact the experts by e-mail and set up a time when you can call them on the phone.

Yes, by phone. E-mail may be convenient for *you*, but most scholars are just

too busy to respond to requests for lists of articles and books. On the other hand, most people *will* talk to you on the phone. A knowledgeable person in the field can give you three or four key citations over the phone right on the spot, and, with the online resources I'm going to tell you about here, that's all you need to get you straight into the literature.

Review Articles

The *Annual Review* series is a good place to start reading. There are *Annual Review* volumes for many disciplines, including psychology (every year since 1950), anthropology (every year since 1972), sociology (since 1975), public health (since 1997), and political science (since 1998). Authors who are invited to publish in these volumes are experts in their fields; they have digested a lot of information and have packaged it in a way that gets you right into the middle of a topic in a hurry. Review articles in journals and bibliographies published as books are two other excellent sources.

Every article in the *Annual Review* series is available online, providing your library subscribes to this service. If it doesn't, then use the printed volumes. Don't worry about printed review articles being out of date. The *Web of Science* and other documentation resources have eliminated the problem of obsolescence in bibliographies and review articles. This will be clear when you read about the *Web of Science* below.

Bibliographic Search Tools: The Online Databases

The overwhelming majority of the research in any discipline is published in thousands of journals (no, I'm not exaggerating), some of which are short lived. A lot of descriptive data on social issues (crime, health care delivery, welfare) is published in reports from governments, industry, and private research foundations. No research project should be launched (and certainly no request for funding of a research project should be submitted) until you have thoroughly searched these potential sources for published research on the topic you are interested in.

Many—but not all—of the bibliographic tools that I describe here are available both online and as paper products. Some of the largest online databases are free, and when that's the case, I provide the Internet address. Many databases are commercial products that are available only by subscription. Check with your school library to see which of these commercial databases they subscribe to. And check with your local city library, also. They may subscribe to

databases that even your college or university doesn't have. Whether or not you use an online service, be sure to use the documentation tools described here when you are starting out on a research project. If online versions are not available, use the paper versions.

A word of caution to new scholars who are writing for publication: Online literature searches make it easy for people to find articles only if the articles (or their abstracts) contain descriptive words. Cute titles on scientific articles hide them from people who want to find them in the indexing tools. If you write an article about illegal Mexican labor migration to the United States and call it something like "Whither Juan? Mexicans on the Road," it's a sure bet to get lost immediately, unless (1) you happen to publish it in one of the most widely read journals and (2) it happens to be a blockbuster piece of work that everyone talks about and is cited in articles that *do* have descriptive titles.

Since most scientific writing is not of the blockbuster variety, you're better off putting words into the titles of your articles that describe what the articles are about. It may seem awfully dull, but descriptive, unimaginative titles are terrific for helping your colleagues find and cite your work.

As formidable as the amount of information being produced in the world is, there is an equally formidable set of documentation resources for accessing information. If you have to use just one of these, then my choice is the *Web of Science*.

Web of Science

The Thompson Institute for Scientific Information, or ISI (<http://www.isinet.com>) produces the *Science Citation Index Expanded* (the SCI), the *Social Sciences Citation Index* (the SSCI), and the *Arts and Humanities Citation Index* (the A&HCI). Together, these three indexes comprise the *Web of Science*, the indispensable resource for doing a literature search. These indexes are available online at most university libraries, and in many small college libraries. I used the paper versions of the SSCI, the SCI, and the A&HCI for 30 years, and if the online versions vanished, I'd go back to the paper ones in a minute. They're that good.

At the ISI, hundreds of people pore over 8,600 journals each year. They examine each article in each journal and enter the title, author, journal, year, and page numbers. In addition, the staff enters all the references cited by each author of each article in each journal surveyed. Some articles have a handful of references, but review articles, like the ones in the *Annual Review* series, can have hundreds of citations. All those citations go into the *Web of Science* databases. So, if you know the name of just one author whose work *should* be cited by anyone working in a particular field, you can find out, for any given

year, who cited that author, and where. In other words, you can search the literature *forward* in time, and this means that older bibliographies, like those in the *Annual Review* series, are never out of date.

Suppose you are interested in the sociolinguistics of African American Vernacular English, also known as Black English. If you ask anyone who has worked on this topic (a sociolinguist in your department, for example) you'll run right into William Labov's *Language in the Inner City: Studies in the Black English Vernacular*, published in 1972. Look up Black English in your library's online catalog and you'll also find *Spoken Soul: The Story of Black English*, by John and Russell Rickford (2000), and *Black Street Speech*, by John Baugh (1983). You'll find Labov's book mentioned prominently in both of the latter books, so right away, you know that Labov's book is a pioneering work and is going to be mentioned by scholars who come later to the topic.

If you are starting a search today of the literature about Black English, you'd want to know who has cited Labov's work, as well as the works of Baugh and the Rickfords. That's where the *Web of Science* comes in. All together, the SSCI, the SCI, and the A&HCI index over 9,000 journals. (The SCI indexes about 6,400 science journals; the SSCI indexes about 1,800 social science journals; and the A&HCI indexes about 1,100 humanities journals.) The SSCI alone indexes about 150,000 articles a year. Ok, so 150,000 *sources* is only a good-sized fraction of the social science papers published in the world each year, but the *authors* of those articles read—and cited—over 2.5 *million* citations to references to the literature. That's 2.5 million citations every year, for decades and decades. Now *that's* a database.

That classic book by Labov on Black English? As of January 2005, some 1,265 researcher articles had cited the book across the sciences and the humanities. The *Web of Science* shows you an abstract for each of those 1,265 articles and lets you examine with a click the entire citation list for each article. You can take any of *those* citations and clip it into another search window and keep the search going. And going, and going.

Don't get overwhelmed by this. You'll be surprised at how quickly you can scan through a thousand hits and pick out things you really need to take a closer look at. As you read some articles and books, you'll find yourself running into the same key literature again and again. That's when you know you're getting to the edges of the world on a particular topic. That's what you want when you're doing research—you want to be working at the edges in order to push the frontiers of knowledge back a bit.

I use the *Web of Science* regularly to keep current with the literature on several topics. I've studied bilingualism in Mexico, for example, particularly the development of bilingual education among the Nāhñu of the Mezquital Valley. The Nāhñu are widely known in the literature as the Otomí, so I scan

the *Web of Science* and look for everything with the word “Otomi” in the title or abstract. Doing this even once a year makes a big difference in your ability to keep up with the expanding literature in every field.

Other Documentation Databases

These days, documentation is big business, and there are many indexing and abstracting resources. Besides the *Web of Science*, three very important resources for anthropologists are *Anthropological Index Online* (AIO), *Anthropological Literature* (AL), and *Abstracts in Anthropology* (AIA). Other useful databases include: *International Bibliography of Anthropology*, ERIC, NTIS, MEDLINE, PsycINFO, *Sociological Abstracts*, Lexis-Nexis, and OCLC.

Anthropological Index Online

Anthropological Index (AI) began as a quarterly journal, published by the Royal Anthropological Institute in London, and as an index of the periodicals in Museum of Mankind Library of the British Museum. AI became an online product in 1997 and is a free service. You’ll find AI-Online at <http://aio.anthropology.org.uk/aio/AIO.html>.

AI covers the major journals for cultural anthropology and archeology, but it also covers many journals that are not covered by any other indexing service, especially small journals from developing nations and eastern Europe. You’ll find articles in AI-Online from journals like the *Annals of the Náprstek Museum* in Prague, the *Quarterly Journal of the Mythic Society*, in Bangalore, and the *Hong Kong Anthropology Bulletin*.

Anthropological Literature

The Tozzer Library in the Peabody Museum of Archaeology and Ethnology at Harvard University is the largest collection of anthropological literature in the world. Beginning in 1963, Tozzer began publishing its card catalog in a set of 52 huge volumes (if you want to see what precomputer library technology was like, just try to lift two or three of those volumes). At the time, the catalog indexed 275,000 items.

In 1979, Tozzer began publishing *Anthropological Literature*, a quarterly journal that indexes the books and articles that come into the library (much as AI indexes the holdings of the Museum of Mankind Library in London). In addition to the library’s book holdings, *Anthropological Literature* indexes about 850 journals across the whole field of anthropology and in related fields

like demography, economics, and psychology. The database for AL grows by about 10,000 citations every year. AL is particularly good for finding older materials on North American, Middle American, and South American archeology and ethnology. The Tozzer Library was founded in 1866, and many of the periodicals received by the library have been indexed since before World War I. You can use AL, then, as a complete index to major journals, such as the *American Anthropologist*, *American Antiquity*, and the like.

AL and AIO are available in one combined database, called Anthropology Plus, through the RLG (Research Libraries Group) Eureka database. If your library doesn't subscribe to this database, you can use AIO at no cost.

Abstracts in Anthropology

AIA is a quarterly journal, published since 1970 (in print form only, not online), that selectively covers current literature on archeology, cultural anthropology, physical anthropology, and linguistics. Indexing journals, like AIO and AL, simply list all the items and cross-index them by author, title, and subject heading. An abstracting journal summarizes the articles it covers by publishing abstracts of from 50 to 200 words.

Indexing services cover more ground; abstracting services provide more depth. AIA publishes 150-word abstracts of the research articles in each of about 130 journals in each issue. AIA publishes the abstracts to all the research articles in the seven most important journals for cultural anthropologists, so browsing through AIA from time to time is a great way to keep up with what's going on in anthropology. The seven top journals are, in alphabetical order: *American Anthropologist*, *American Ethnologist*, *Current Anthropology*, *Ethnology*, *Human Organization*, *Journal of Anthropological Research*, and the *Journal of the Royal Anthropological Institute*.

AIA covers some journals not covered by other publications—journals like *Oral History* (published by the Institute of Papua New Guinea) and *Caribbean Studies* (published by the Institute of Caribbean Studies at the University of Puerto Rico). The SSCI didn't cover the *Papers in Anthropology* series of the University of Oklahoma, but AIA used to cover it, and one of the papers abstracted by AIA in 1983 was by G. Agogino and B. Ferguson on an Indian-Jewish community in the state of Hidalgo, Mexico, very close to the Nāhñu Indian communities I've studied.

Of course, I would have located the paper through the SSCI had anyone cited it in one of the three million articles that the SSCI has indexed just since 1984 (the first year the article could have been cited), but a check revealed that no one did cite it, so looking through AIA was probably the only way I could have run into that particular piece of work.

International Bibliography of Anthropology

This source is part of the *International Bibliography of the Social Sciences* (IBSS) series that began in 1952 as a product of the International Committee on Social Science Information and Documentation (ICSSID), a UNESCO-funded body. There were four volumes in the set each year, one each on sociology, political science, economics, and cultural anthropology. In recent years, the IBSS has been published commercially and is now both a paper and an online product. The unique thing about the IBSS is that the volumes are based on data submitted by librarians around the world (from Thailand, Haiti, Zambia, Hungary, Argentina, etc.) who document the social science information being produced in their countries.

This information is entered into a computer and is sorted and selected for inclusion in each year's volumes. The procedure takes a long time, so the paper bibliographies are not all that current, but they are a very good source for locating materials published by national and regional journals in the developing world and in eastern European countries. The online version of the IBSS, however, is updated weekly, and it contains all the information from over 50 years of documentation effort.

ERIC

The Educational Resources Information Center, or ERIC, began in 1966 as a microfiche archive to index literature of interest to researchers in education. ERIC is free and online at <http://www.eric.ed.gov>. Many of the journals that ERIC indexes are of direct relevance to work in anthropology, and with over a million citations, it's a treasure. But the unique advantage of ERIC is the access it gives you to the gray literature—over 100,000, full-text research reports on studies funded by government agencies and by private foundations. (The database is updated weekly and grows by about 34,000 documents every year.)

For example, I follow the efforts of indigenous peoples around the world to keep their languages alive. I found a report in ERIC by Marion Blue Arm, published in a conference proceeding, on attitudes by Indians, Whites, and mixed families toward the teaching of Lakota in the schools on the Cheyenne River Sioux Reservation in South Dakota. ERIC is filled with useful material like that.

NTIS

NTIS, the National Technical Information Service, indexes and abstracts federally funded research reports in all areas of science. Many technical

reports eventually get published as articles, which means you can find them through all the other databases. But many research reports just get filed and shelved—and lost. A major reason that technical reports don't get published is that they contain data—huge tables of the stuff—which journals don't have room for.

The NTIS has technical reports from archeological digs, from focus groups on attitudes about unprotected sex, from development projects on giant clam mariculture in the Solomon Islands, from natural experiments to test how long people can stay in a submerged submarine without going crazy—if the federal government has funded it under contract, there's probably a technical report of it. The NTIS is available online at <http://www.ntis.gov/>. Papers written since 1990 are available in microfiche, for a fee.

MEDLINE

MEDLINE is a product of the National Library of Medicine. It covers over 3,700 journals in the medical sciences—including the medical social sciences—going back to 1966. MEDLINE is available at university libraries through services like FirstSearch and Cambridge Scientific Abstracts, but there is a free version of MEDLINE, called, whimsically, *PubMed*, at <http://www.index.nlm.nih.gov/databases/freemedl.html>.

If you are working on anything that has to do with health care, MEDLINE, with its four million citations and abstracts is a must.

PsycINFO, PsycARTICLES, and Sociological Abstracts

PsycINFO and PsycARTICLES are products of the American Psychological Association. The Jurassic version of PsycINFO goes back to 1887 and has over eight million citations. The database covers 1,800 journals (including the *American Anthropologist*) and adds about 50,000 new references and abstracts each year. The PsycARTICLES database comprises full-text material from journals published since 1985 by the American Psychological Association and by the Educational Publishing Foundation.

Sociological Abstracts is a product of Sociological Abstracts, Inc., a division of Cambridge Scientific Abstracts. It covers about 1,800 journals dating from 1963. SA contains about 650,000 records, with about 28,000 new records added each year.

PsycINFO and *Sociological Abstracts* both have excellent coverage of research methods, the sociology of language, occupations and professions, health, family violence, poverty, and social control. They cover the sociology

of knowledge and the sociology of science, as well as the sociology of the arts, religion, and education.

Linguistics and Language Behavior Abstracts

LLBA, published by Sociological Abstracts, Inc., indexes and abstracts journals in descriptive linguistics, sociolinguistics, anthropological linguistics, psycholinguistics, and so on. It is an excellent resource for work on discourse analysis and all forms of text analysis. The database contained about 360,000 entries in 2005 and is growing at about 18,000 entries a year.

LEXIS/NEXIS

LEXIS/NEXIS (<http://www.lexis.com/>) began in 1973 as a way to help lawyers find information on cases. Today, the Lexis-Nexis Universe database contains the actual text of about ten million articles from over 18,000 sources, including more than 5,600 news, business, legal, and medical sources. This source includes 55 of the world's major newspapers and 65 newspapers and magazines that target specific ethnic groups in the United States. If your library gives you access to LEXIS-NEXIS, then no literature search is complete until you've used it. LN has the complete *New York Times* from June 1, 1980, to the present. For more historical documents (especially for historical data on ethnic groups in the United States), use the *New York Times* full-text database, if your library has it. That index goes back to the first issue on Sept. 18, 1851.

The CIS

Lexis-Nexis has grown by acquiring several other major databases. One is the CIS, or Congressional Information Service. This service indexes U.S. House and Senate hearings, reports entered into public access by submission to Congress, and testimony before congressional committees. All of these, are in print and available to the public, free, either through libraries (at least one library in every state in the United States is designated a Federal Depository Library and has every document printed by the U.S. government) or through the U.S. Government Printing Office.

You can get access to any public documents published by the U.S. Congress at <http://thomas.loc.gov/> (the "thomas" refers to Thomas Jefferson), but if you have access to Lexis-Nexis, it's easier to use that service to find things in the CIS. There are congressional reports on many topics of interest to anthropolo-

gists, including reports on current social issues (housing, nutrition, cultural conservation, rural transportation). The proceedings for recognizing American Indian tribes are published in the *Congressional Record* and are available through CIS, as are reports on the demographics of American ethnic groups.

OCLC

OCLC (<http://www.oclc.org/oclc/menu/home1.htm>), the Online Computer Library Center, is the world's largest library database. In 2005, OCLC covered 15,000 journals and had bibliographic records for over 56 million books—and was growing fast. If you find a book or article in the *Web of Science* or PsycINFO, etc., and your library doesn't have it, then the OCLC will tell you which library *does* have it. Interlibrary loans depend on OCLC. The components of OCLC include WorldCat (the world catalog of books) for books and Article First (for journals).

Some Additional Websites

In addition to the usual library documentation resources, many super information resources of interest to anthropologists are available through the Internet pages of international organizations. A good place to start is the University of Michigan Document Center's page, called "International Agencies and Information on the Web" at <http://www.lib.umich.edu/govdocs/intl.html>. This page will point you to the big meta-sites that list the websites for thousands of intergovernmental and nongovernmental organizations (widely known as IGOs and NGOs). The Union of International Associations' guide to IGOs and NGOs, for example, lists thousands of sites. It's at <http://www.uia.org/website.htm>.

The Michigan Documents Center site will also point you to dozens of international agency sites, places like the Food and Agriculture Organization of the UN, with many full-text reports on its projects since 1986. Interested in "shrimp aquaculture"? Go to FAO's archiving site at <http://www.fao.org/documents/> and type that phrase into the search engine to find reports on local farming practices. Go to the site of the United Nations High Commission on Refugees (UNHCR) at <http://www.unhcr.ch/> for the latest statistics on the refugees around the world. Go to the site of the Pan American Health Organization (PAHO) at <http://www.paho.org/> for the latest statistics on health indicators for countries in North and South America.

There is a world of high-quality documents available on the Internet. No search of the literature is now complete without combing through those resources. The Anthropology Review Database at <http://wings.buffalo.edu/ARD/>

provides an expanding list of reviews of books, articles, and software of interest to anthropologists. The Scout Report for Social Sciences, at <http://scout.cs.wisc.edu/>, is an expanding database of Internet sites. It's published weekly. Search the site for "anthropology" and you'll find lots of useful, full-text resources, like the one at <http://www.nativetech.org/>, devoted to Native American technology and art.

To find general information about particular countries of the world, start with the Yahoo server <http://dir.yahoo.com/Regional/Countries/> and also consult the *CIA Factbook* at <http://www.odci.gov/cia/publications/factbook/index.html>. Going to the Manu'a Islands to do fieldwork? Detailed maps for just about every country and region of the world (including the Manu'a Islands) are available online at <http://www.lib.utexas.edu/maps/index.html> at the University of Texas. For websites developed and maintained by particular countries, start with <http://www.library.nwu.edu/govpub/resource/internat/foreign.html> at Northwestern University. All government statistical reports are subject to error. Use with caution.

Tips for Searching Online Databases

1. Make sure you get the spelling right. It's not as easy as it sounds. If you ask the search engine at Yahoo.com or MSN.com for information on "New Guinae," it'll ask if you really meant "New Guinea." Unfortunately, as of 2005, many of the databases I've told you about lack such intelligent spell-checkers, so if you make a mistake, you're on your own. Keep a list of commonly misspelled words handy on your computer (misspelled is commonly misspelled, of course) so if you ask for references on "apropriate technology" and it comes back, incongruously, with "nothing found," you can figure out what's wrong. Or type any word you're not sure about into <http://dictionary.reference.com/>. If you've got the wrong spelling, that site's intelligent spell-checker will try to figure out what you meant to say.
2. If there are two ways to spell a word, be sure to search with both spellings. Use both Koran and Qur'an (and several other spelling variants) in your searches; Chanukah and Hanukah (and several other spelling variants); Rom and Roma (both are used to refer to Gypsies); Thessaloniki and Salonika; Madras and Chennai; Beijing and Peking; and so on.

I asked MEDLINE for articles from 1990 until January 2005 on (childhood diarrhea) OR (infantile diarrhea) and got 2,730. Then I changed the spelling to (childhood diarrhoea) OR (infantile diarrhoea)—the spelling more commonly used in Britain—and got *another* 1,056 hits.

I searched *Sociological Abstracts* for references to the Attitudes Toward

Women Scale and turned up 50 references. I found another four references by changing “toward” to “towards.” Now, “towards” was simply not the word in the original name of the scale, but after more than 30 years of use, I was betting that someone would have used that variant.

Some databases do have intelligent spell-checkers. A search for “behavior measurement” and “behaviour measurement” in PsycINFO turned up the same list of 142 references. But neither spelling captured the 230 references that I got with “behavioral measurement” or the nine additional references for “behavioural measurement.” I had a similar experience with “geographic information systems” and “geographical information systems.” These quirks in the search engines are simply there, so learn to search creatively.

3. Learn to narrow your searches. This just takes practice. That MEDLINE search for (childhood diarrhea) OR (infantile diarrhea) and for (childhood diarrhoea) OR (infantile diarrhoea) turned up a total of 3,786 articles. I narrowed the search to (childhood diarrhea OR infantile diarrhea) AND (cultural factors), and that returned 29 items. Changing “diarrhea” to “diarrhoea” in that search added 15 more items.

The words AND and OR in these searches are called **Boolean operators**. The other operator is NOT. I asked MEDLINE for all articles since 1977 on (regimen compliance AND malaria). That search brought back abstracts for 54 articles. When I restricted the search to [(regimen compliance AND malaria) NOT Africa], I got back 37 abstracts. The parentheses and brackets in Boolean searches work just like equations in algebra. So, that last search I did fetched all the items that had “regimen compliance” and “malaria” in their titles or abstract. Then, the items were sorted and any of those items that contained the word “Africa” was dropped. The logic of Boolean operators is used in database management—about which, lots more at the end of chapter 14 on field notes.

Meta-Analysis

Meta-analysis involves piling up all the quantitative studies ever done on a particular topic to assess quantitatively what is known about the size of the effect. The pioneering work on meta-analysis (M. L. Smith and Glass 1977) addressed the question: Does psychotherapy make a difference? That is, do people who get psychotherapy benefit, compared to people who have the same problems and who don’t get psychotherapy? Since then, there have been thousands of meta-analyses, but most of them are in fields like psychology and medicine, where data from experiments lend themselves to direct comparison.

A few anthropologists, though, have done important meta-analyses. Until the 1970s, conventional wisdom had it that hunters and gatherers lived a precarious existence, searching all the time for food, never knowing where their next meal was coming from. In 1970, Esther Boserup made the important observation that plow agriculture takes more time than does hoe agriculture. And in 1972, Marshall Sahlins generalized this observation, arguing that hunter-gatherer societies had more leisure than anyone else and that people have to work more and more as societies become more complex.

In 1996, Ross Sackett did a meta-analysis of 102 cases of time allocation studies and 207 energy-expenditure studies to test Sahlins's (1972) primitive affluence hypothesis. Sackett found that, in fact, adults in foraging and horticultural societies work, on average, about 6.5 hours a day, while people in agricultural and industrial societies work about 8.8 hours a day. The difference, by the way, is statistically very significant (Sackett 1996:231, 547).

Meta-analysis can be delightfully subversive. Morrison and Morrison (1995), for example, found that only 6.3% of the variance in graduate-level GPA is predicted by performance on the GRE quantitative and verbal exams. And White (1980) found that across a hundred studies up to 1979, socioeconomic status explained, on average, an identical 6.3% of the variance in school achievement. The raw correlation across those hundred studies, ranged from $-.14$ (yes, *minus* .14) to .97.

I think that meta-analyses will become more and more popular, as electronic databases, including databases of ethnography, develop. Meta-analysis gives us the tools to see if we are making progress on important question. It is, as Hunt (1997) says, the way science takes stock.

5



Research Design: Experiments and Experimental Thinking

Early in the 20th century, F. C. Bartlett went to Cambridge University to study with W. H. R. Rivers, an experimental psychologist. In 1899, Rivers had been invited to join the Torres Straits Expedition and saw the opportunity to do comparative psychology studies of non-Western people (Tooker 1997:xiv). When Bartlett got to Cambridge, he asked Rivers for some advice. Bartlett expected a quick lecture on how to go out and stay out, about the rigors of fieldwork, and so on. Instead, Rivers told him: “The best training you can possibly have is a thorough drilling in the experimental methods of the psychological laboratory” (Bartlett 1937:416).

Bartlett found himself spending hours in the lab, “lifting weights, judging the brightness of lights, learning nonsense syllables, and engaging in a number of similarly abstract occupations” that seemed to be “particularly distant from the lives of normal human beings.” In the end, though, Bartlett concluded that Rivers was right. Training in the **experimental method**, said Bartlett, gives one “a sense of evidence, a realization of the difficulties of human observation, and a kind of scientific conscience which no other field of study can impart so well” (ibid.:417).

I agree. Most anthropologists don’t do experiments, but a solid grounding in the **logic of experiments** is one of the keys to good research skills, no matter what kind of research you’re doing.

At the end of this chapter, you should understand the variety of **research designs**. You should also understand the concept of **threats to validity** and how we can respond to those threats.

Experiments

There are two ways to categorize experiments. First, in **true experiments**, participants are assigned randomly, to either a **treatment group** or a **control group**, while in **quasi-experiments**, participants are selected rather than assigned. (By the way, I prefer “participants” to “subjects” when we’re talking about people who take part in experiments.)

Second, in **laboratory experiments** you have the luxury of controlling variables, while in **field experiments** you get far greater realism. The logic of experiments is the same no matter where they’re done. There are, of course, differences in experiments with people vs. experiments with rocks or pigeons or plants. But these differences involve ethical issues—like deception, informed consent, and withholding of treatment—not logic. More on these ethical issues later.

True Experiments

There are five steps in a **classic experiment**:

1. Formulate a hypothesis.
2. Randomly assign participants to the intervention group or to the control group.
3. Measure the dependent variable(s) in one or both groups. This is called O_1 or “observation at time 1.”
4. Introduce the treatment or intervention.
5. Measure the dependent variable(s) again. This is called O_2 or “observation at time 2.”

Later, I’ll walk you through some variations on this five-step formula, including one very important variation that does not involve Step 3 at all. But first, the basics.

Step 1.

Before you can do an experiment, you need a clear hypothesis about the relation between some independent variable (or variables) and some dependent variable (or variables). Experiments thus tend to be based on **confirmatory** rather than **exploratory research** questions.

The testing of new drugs can be a simple case of one independent and one dependent variable. The independent variable might be, say, “taking vs. not taking” a drug. The dependent variable might be “getting better vs. not getting better.” The independent and dependent variables can be much more subtle. “Taking vs. not taking” a drug might be “taking more of, or less of” a drug,

and “getting better vs. not getting better” might be “the level of improvement in high-density lipoprotein” (the so-called good cholesterol).

Move this logic to agriculture: *ceteris paribus* (holding everything else—like amount of sunlight, amount of water, amount of weeding—constant), some corn plants get a new fertilizer and some don’t. Then, the dependent variable might be the number of ears per corn stalk or the number of days it takes for the cobs to mature, or the number of kernels per cob.

Now move this same logic to human thought and human behavior: *Ceteris paribus*, people in Nairobi who take this course in AIDS awareness will report fewer high-risk sex practices than will people who don’t take this course. *Ceteris paribus* here means that people in both groups—the treatment group and the control group—start with the same amount of reported high-risk sexual activity.

Things get more complicated, certainly, when there are multiple independent (or dependent) variables. You might want to test two different courses, with different content, on people who come from three different tribal backgrounds. But the underlying logic for setting up experiments and for analyzing the results is the same across the sciences. When it comes to experiments, everything starts with a clear hypothesis.

Step 2.

You need at least two groups, called the **treatment group** (or the **intervention group** or the **stimulus group**) and the **control group**. One group gets the intervention (the new drug, the new teaching program) and the other group doesn’t. The treatment group (or groups) and the control group(s) are involved in different **experimental conditions**.

In true experiments, people are **randomly assigned** to either the intervention group or to the control group. This ensures that any differences between the groups is the consequence of chance and not of **systematic bias**. Some people in a population may be more religious, or more wealthy, or less sickly, or more prejudiced than others, but random assignment ensures that those traits are randomly distributed through all the groups in an experiment.

Random assignment doesn’t eliminate selection bias. It makes differences between experimental conditions (groups) due solely to chance by taking the decision of who goes in what group out of your hands. The principle behind random assignment will become clearer after you work through chapter 6 on probability sampling, but the bottom line is this: Whenever you *can* assign participants randomly in an experiment, do it.

Step 3.

One or both groups are measured on one or more dependent variables. This is called the **pretest**.

Dependent variables in people can be physical things like weight, height, systolic blood pressure, or resistance to malaria. They can also be attitudes, moods, knowledge, or mental and physical achievements. For example, in weight-loss programs, you might measure the ratio of body fat to body mass as the dependent variable. If you are trying to raise women's understanding of the benefits of breast-feeding by exposing them to a multimedia presentation on this topic, then a preliminary test of women's attitudes about breast-feeding before they see the presentation is an appropriate pretest for your experiment.

You don't always need a pretest. More on this in a bit, when we discuss threats to validity in experiments.

Step 4.

The intervention (the independent variable) is introduced.

Step 5.

The dependent variables are measured again. This is the **posttest**.

A Walkthrough

Here's a made-up example of a true experiment: Take 100 college women (18–22 years of age) and randomly assign them to two groups of 50 each. Bring each woman to the lab and show her a series of flash cards. Let each card contain a single, three-digit random number. Measure how many three-digit numbers each woman can remember. Repeat the task, but let the members of one group hear the most popular rock song of the week playing in the background as they take the test. Let the other group hear nothing. Measure how many three-digit numbers people can remember and whether rock music improves or worsens performance on the task.

Do you think this is a frivolous experiment? Many college students study while listening to rock music, which drives their parents crazy. I'll bet that more than one reader of this book has been asked something like: "How can you learn anything with all that noise?" The experiment I've outlined here is designed to test whether students can, in fact, "learn anything with all that noise."

Of course, this experiment is very limited. Only women are involved. There are no graduate students or high school students. There's no test of whether classic rock helps or hinders learning more than, say, rhythm and blues, or country music, or Beethoven. And the learning task is artificial. What we know at the end of this experiment is whether college-age women learn to memorize more or fewer three-digit numbers when the learning is accompanied by a single rock tune.

But a lot of what's really powerful about the experimental method is embodied in this example. Suppose that the rock-music group does better on the task. We can be pretty sure this outcome is not because of the participants' sex, age, or education, but because of the music. Just sticking in more independent variables (like expanding the group to include men, graduate students, or high school students; or playing different tunes; or making the learning task more realistic), without modifying the experiment's design to control for all those variables, creates what are called **confounds**. They *confound* the experiment and make it impossible to tell if the intervention is what really caused any observed differences in the dependent variable.

Good experiments test narrowly defined questions. This is what gives them knowledge-making power. When you do a good experiment, you *know* something at the end of it. In this case, you know that women students at one school memorize or do not memorize three-digit numbers better when they listen to a particular rock tune.

This may not seem like much, but you really know it. You can repeat the experiment at the same school to verify or refute this little bit of knowledge. You can repeat the experiment at another school to see if the knowledge has external validity.

Suppose you don't get the same answer at another school, holding all the other elements of the experiment—age, sex, type of music—constant. The new finding demands an explanation. Perhaps there is something about the student selection process at the two schools that produces the different results? Perhaps students at one school come primarily from working-class families, while students from the other school come from upper-middle-class families. Perhaps students from different socioeconomic classes grow up with different study habits, or prefer different kinds of music.

Conduct the experiment again but include men this time. Conduct it again and include two music conditions: a rock tune and a classical piece. Take the experiment on the road and run it all over again at different-sized schools in different regions of the country. Then, on to Paraguay. . . .

True experiments, with randomized assignment and full control by the researcher, produce knowledge that has high **internal validity**. This means that changes in the dependent variables were probably *caused by*—not merely related to or correlated with—the treatment. Continual replication produces **cumulative knowledge** with high **external validity**—that is, knowledge that you can generalize to people who were not part of your experiment.

Replication of knowledge is every bit as important as its production in the first place. In fact, in terms of usefulness, replicated knowledge is exactly what we're after.

Kinds of Confounds: Threats to Validity

It's pointless to ask questions about external validity until you establish internal validity. In a series of influential publications, Donald Campbell and his colleagues identified the threats to internal validity of experiments (see Campbell 1957, 1979; Campbell and Stanley 1966; Cook and Campbell 1979). Here are seven of the most important confounds:

1. *History*

The **history confound** refers to any independent variable, other than the treatment, that (1) occurs between the pretest and the posttest in an experiment and (2) affects the experimental groups differently. Suppose you are doing a laboratory experiment, with two groups (experimental and control) and there is a power failure in the building. So long as the lights go out for both groups, there is no problem. But if the lights go out for one group and not the other, it's difficult to tell whether it was the treatment or the power failure that causes changes in the dependent variable.

In a laboratory experiment, history is controlled by isolating participants as much as possible from outside influences. When we do experiments outside the laboratory, it is almost impossible to keep new independent variables from creeping in and confounding things.

Here's an example of an experiment outside the lab. Suppose you run an experiment to test whether monetary incentives help third graders do better in arithmetic. Kids in the treatment classes get a penny for each right answer on their tests; kids in the control classes get nothing. Now, right in the middle of the school term, while you're running this experiment, the Governor's Task Force on Education issues its long-awaited report, with a recommendation that arithmetic skills be emphasized during the early school years. Furthermore, it says, teachers whose classes make exceptional progress in this area should be rewarded with 10% salary bonuses.

The governor accepts the recommendation and announces a request for a special legislative appropriation. Elementary teachers all over the state start paying extra attention to arithmetic skills. Even supposing that the students in the treatment classes do better than those in the control classes in your experiment, we can't tell if the magnitude of the difference would have been greater had this historical confound not occurred. That's just the breaks of experimenting in real life without being able to control everything.

2. *Maturation*

The **maturation confound** refers to the fact that people in any experiment grow older or get more experienced while you are trying to conduct an experi-

ment. Consider the following experiment: Start with a group of teenagers on a Native American reservation and follow them for the next 60 years. Some of them will move to cities, some will go to small towns, and some will stay on the reservation. Periodically, test them on a variety of dependent variables (their political opinions, their wealth, their health, their family size, and so on). See how the experimental treatments (city vs. reservation vs. town living) affect these variables.

Here is where the maturation confound enters the picture. The people you are studying get older. Older people in many societies become more politically conservative. They are usually wealthier than younger people. Eventually, they come to be more illness prone than younger people. Some of the changes you measure in your dependent variables will be the result of the various treatments—and some of them may just be the result of maturation.

Maturation is not just about people getting older. Social service delivery programs “mature” by working out bugs in their administration. People “mature” through practice with experimental conditions and they become fatigued. We see this all the time in new social programs where people start out being really enthusiastic about innovations in organizations and eventually get bored or disenchanting.

3. *Testing and Instrumentation*

The **testing confound** happens when people change their responses in reaction to being constantly examined. Asking people the same questions again and again in a longitudinal study, or even in an ethnographic study done over 6 months or more, can have this effect.

The **instrumentation confound** results from changing measurement instruments. Changing the wording of questions in a survey is essentially changing instruments. Which responses do you trust: the ones to the earlier wording or the ones to the later wording? If you do a set of observations in the field and later send in someone else to continue the observations you have changed instruments. Which observations do you trust as closer to the truth: yours or those of the substitute instrument (the new field researcher)?

In multi-researcher projects, this problem is usually dealt with by training all investigators to see and record things in more or less the same way. This is called increasing **interrater reliability**. (More on this in chapter 17, on text analysis.)

4. *Regression to the Mean*

Regression to the mean is a confound that can occur when you study groups that have extreme scores on a dependent variable. No matter what the

intervention is, the extreme scores are likely to become more moderate just because there's nowhere else for them to go.

If men who are taller than 6'7" marry women who are taller than 6'3", then their children are likely to be (1) taller than average and (2) closer to average height than either of their parents are. There are two independent variables (the height of each of the parents) and one dependent variable (the height of the children). We expect the dependent variable to "regress toward the mean," since it really can't get more extreme than the height of the parents.

I put that phrase "regress toward the mean" in quotes because it's easy to misinterpret this phenomenon—to think that the "regressing" toward the mean of an dependent variable is caused by the extreme scores on the independent variables. It isn't, and here's how you can tell that it isn't: Very, very tall children are likely to have parents whose height is more like the mean. One thing we know for sure is that the height of children doesn't cause the height of their parents. Regression to the mean is a statistical phenomenon—it happens in the aggregate and is not something that happens to individuals.

Many social intervention programs make the mistake of using people with extreme values on dependent variables as subjects. Run some irrigation canals through the most destitute villages in a region of a Third World country and watch the average income of villagers rise. But understand that income might have risen anyway, if you'd done nothing, because it couldn't have gone down. Test a reading program on the kids in a school district who score in the bottom 10% of all kids on reading skills and watch their test scores rise. But understand that their scores might have risen anyway.

5. Selection of Participants

Selection bias in choosing participants is a major confound to validity. In true experiments, you assign participants at random, from a single population, to treatment groups and control groups. This distributes any differences among individuals in the population throughout the groups, making the groups equivalent.

This reduces the possibility that differences among the groups will cause differences in outcomes on the dependent variables. Random assignment in true experiments, in other words, maximizes the chance for valid outcomes—outcomes that are not clobbered by hidden factors.

In natural experiments, however, we have *no control* over assignment of individuals to groups.

Question: Do victims of violent crime have less stable marriages than persons who have not been victims? Obviously, researchers cannot randomly assign participants to the treatment (violent crime). It could turn out that peo-

ple who are victims of this treatment are more likely to have unstable marriages anyway, even if they never experience violence.

Question: Do migrants to cities from villages in developing nations engage in more entrepreneurial activities than stay-at-homes? If we could assign rural people randomly to the treatment group (those engaging in urban migration), we'd have a better chance of finding out. But we can't, so selection is a threat to the internal validity of the experiment. Suppose that the answer to the question at the top of this paragraph were "yes." We still don't know the direction of the causal arrow: Does the treatment (migration) cause the outcome (greater entrepreneurial activity)? Or does having an entrepreneurial personality cause migration?

6. Mortality

The **mortality confound** refers to the fact that people may not complete their participation in an experiment. Suppose we follow two sets of Mexican villagers—some who receive irrigation and some who do not—for 5 years. During the 1st year of the experiment, we have 200 villagers in each group. By the 5th year, 170 remain in the treatment group, and only 120 remain in the control group. One conclusion is that lack of irrigation caused those in the control group to leave their village at a faster rate than those in the treatment group.

But what about those 30 people in the treatment group who left? It could be that they moved to another community where they acquired even more irrigated land, or they may have abandoned farming altogether to become labor migrants. These two outcomes would affect the results of the experiment quite differently. Mortality can be a serious problem in natural experiments if it gets to be a large fraction of the group(s) under study.

Mortality also affects panel surveys. That's where you interview the same people more than once to track something about their lives. We'll talk more about those in chapter 10.

7. Diffusion of Treatments

The **diffusion of treatments** threat to validity occurs when a control group cannot be prevented from receiving the treatment in an experiment. This is particularly likely in quasi-experiments where the independent variable is an information program.

In a project with which I was associated some time ago, a group of African Americans were given instruction on modifying their diet and exercise behavior to lower their blood pressure. Another group was randomly assigned from

the population to act as controls—that is, they would not receive instruction. The evaluation team measured blood pressure in the treatment group and in the control group before the program was implemented. But when they went back after the program was completed, they found that control group members had also been changing their behavior. They had learned of the new diet and exercises from the members of the treatment group.

Controlling for Threats to Validity

In what follows, I want to show you how the power of experimental logic is applied to real research problems. The major experimental designs are shown in figure 5.1. The notation is pretty standard. X stands for some intervention—a stimulus or a treatment. R means that participants are randomly assigned to experimental conditions—either to the intervention group that gets the treatment, or to the control group that doesn't. Several designs include random assignment and several don't. O stands for "observation." O_1 means that some observation is made at time 1, O_2 means that some observation is made at time 2, and so on.

"Observation" means "measurement of some dependent variable," but as you already know, the idea of measurement is pretty broad. It can be taking someone's temperature or testing their reading skill. It can also be just writing down whether they were successful at hunting game that day.

The Classic Two-Group Pretest-Posttest Design with Random Assignment

Figure 5.1a shows the classic experimental design: the **two-group pretest-posttest** with random assignment. From a population of potential participants, some participants have been assigned randomly to a treatment group and a control group. Read across the top row of the figure. An observation (measurement) of some dependent variable or variables is made at time 1 on the members of group 1. That is O_1 . Then an intervention is made (the group is exposed to some treatment, X). Then, another observation is made at time 2. That is O_2 .

Now look at the second row of figure 5.1a. A second group of people are observed, also at time 1. Measurements are made of the same dependent variable(s) that were made for the first group. The observation is labeled O_3 . There is no X on this row, which means that no intervention is made on this group of people. They remain unexposed to the treatment or intervention in the

	Time 1		Time 2	
	Assignment	Pretest	Intervention	Posttest
Group 1	R	O ₁	X	O ₂
Group 2	R	O ₃		O ₄
a. The Classic Design: Two-Group Pretest-Posttest				
Group 1	R	O ₁	X	O ₂
Group 2	R	O ₃		O ₄
Group 3	R		X	O ₅
Group 4	R			O ₆
b. The Solomon Four-Group Design				
Group 1		O ₁	X	O ₂
Group 2		O ₃		O ₄
c. The Classic Design Without Randomization				
Group 1	R		X	O ₁
Group 2	R			O ₂
d. The Campbell and Stanley Posttest-Only Design				
			X	O
e. The One-Shot Case Study Design				
		O ₁	X	O ₂
f. The One-Group Pretest-Posttest Design				
			X	O ₁
				O ₂
g. Two-Group Posttest Only: Static Group Comparison Design				
		O O O	X	O O O
h. The Interrupted Time Series Design				

Figure 5.1. Some research designs.

experiment. Later, at time 2, after the first group has been exposed to the intervention, the second group is observed again. That’s O₄.

Random assignment of participants ensures equivalent groups, and the second group, without the intervention, ensures that several threats to internal validity are taken care of. Most importantly, you can tell how often (how many times out of a hundred, for example) any differences between the pretest

	<i>Time 1</i>		<i>Time 2</i>	
	<i>Assignment</i>	<i>Pretest</i>	<i>Intervention</i>	<i>Posttest</i>
Group 1	<i>R</i>	O_1	<i>X</i>	O_2
Group 2	<i>R</i>	O_3		O_4

Figure 5.1a. The classic design: Two-group pretest-posttest.

and posttest scores for the first group might have occurred anyway, even if the intervention hadn't taken place.

Patricia Chapman (Chapman et al. 1997) wanted to educate young female athletes about sports nutrition. She and her colleagues worked with an eight-team girl's high school softball league in California. There were nine 14–18 years olds on each team, and Chapman et al. assigned each of the 72 players randomly to one of two groups. In the treatment group, the girls got two, 45-minute lectures a week for 6 weeks about dehydration, weight loss, vitamin and mineral supplements, energy sources, and so on. The control group got no instruction.

Before the program started, Chapman et al. asked each participant to complete the Nutrition Knowledge and Attitude Questionnaire (Werblow et al. 1978) and to list the foods they'd eaten in the previous 24 hours. The nutrition knowledge-attitude test and the 24-hour dietary recall test were the **pretests** in this experiment. Six weeks later, when the program was over, Chapman et al. gave the participants the same two tests. These were the **posttests**. By comparing the data from the pretests and the posttests, Chapman et al. hoped to test whether the nutrition education program had made a difference.

The education intervention did make a difference—in knowledge, but not in reported behavior. Both groups scored about the same in the pretest on knowledge and attitudes about nutrition, but the girls who went through the lecture series scored about 18 points more (out of 200 possible points) in the posttest than did those in the control group.

However, the program had no effect on what the girls reported eating. After 6 weeks of lectures, the girls in the treatment group reported consuming 1,892 calories in the previous 24 hours, while the girls in the control group reported 1,793 calories. A statistical dead heat. This is not nearly enough for young female athletes, and the results confirmed for Chapman what other studies had already shown—that for many adolescent females, the attraction of competitive sports is the possibility of losing weight.

This classic experimental design is used widely to evaluate educational pro-

grams. Kunovich and Rashid (1992) used this design to test their program for training freshman dental students in how to handle a mirror in a patient’s mouth (think about it; it’s not easy—everything you see is backward).

The Solomon Four-Group Design

The classic design has one important flaw—it’s subject to testing bias. Differences between variable measurements at time 1 and time 2 might be the result of the intervention, but they also might be the result of people getting savvy about being watched and measured. Pretesting can, after all, sensitize people to the purpose of an experiment, and this, in turn, can change people’s behavior. The **Solomon four-group design**, shown in figure 5.1b, controls for this. Since there are no measurements at time 1 for groups 3 and 4, this problem is controlled for.

	<i>Time 1</i>		<i>Time 2</i>	
	<i>Assignment</i>	<i>Pretest</i>	<i>Intervention</i>	<i>Posttest</i>
Group 1	R	O ₁	X	O ₂
Group 2	R	O ₃		O ₄
Group 3	R		X	O ₅
Group 4	R			O ₆

Figure 5.1b. The Solomon four-group design.

Larry Leith (1988) used the Solomon four-group design to study a phenomenon known to all sports fans as the “choke.” That’s when an athlete plays well during practice and then loses it during the real game, or plays well all game long and folds in the clutch when it really counts. It’s not pretty.

Leith assigned 20 male students randomly to each of the four conditions in the Solomon four-group design. The pretest and the posttest were the same: Each participant shot 25 free throws on a basketball court. The dependent variable was the number of successful free throws out of 25 shots in the posttest. The independent variable—the treatment—was giving or not giving the following little pep talk to each participant just before he made those 25 free throws for the posttest:

Research has shown that some people have a tendency to choke at the free-throw line when shooting free throws. No one knows why some people tend to choking

behavior. However, don't let that bother you. Go ahead and shoot your free throws. (Leith 1988:61)

What a wonderfully simple, utterly diabolic experiment. You can guess the result: There was a significantly greater probability of choking if you were among the groups that got that little pep talk, irrespective of whether they'd been given the warm-up pretest.

The Two-Group Pretest-Posttest without Random Assignment

Figure 5.1c shows the design for a **quasi-experiment**—an experiment in which participants are *not* assigned randomly to the control and the experi-

	<i>Time 1</i>		<i>Time 2</i>	
	<i>Assignment</i>	<i>Pretest</i>	<i>Intervention</i>	<i>Posttest</i>
Group 1		O_1	X	O_2
Group 2		O_3		O_4

Figure 5.1c. The classic design without randomization: The quasi-experiment.

mental condition. This compromise with design purity is often the best we can do.

Program evaluation research is usually quasi-experimental. Consider a program in rural Kenya in which women farmers are offered instruction on applying for bank credit to buy fertilizer. The idea is to increase corn production.

You select two villages in the district—one that gets the program, and one that doesn't. Before the program starts, you measure the amount of credit, on average, that women in each village have applied for in the past 12 months. A year later, you measure again and find that, on average, women in the program village have applied for more agricultural credit than have their counterparts in the control village.

Campbell and Boruch (1975) show how this research design leads to problems. But suppose that the women in the program village have, on average, more land than the women in the control village have. Would you (or the agency you're working for) be willing to bet, say, \$300,000 on implementing the program across the district, in, say, 30 villages? Would you bet that it was the intervention and not some confound, like the difference in land holdings, that caused the difference in outcome between the two villages?

The way around this is to assign each woman randomly to one of the two

conditions in the experiment. Then, the confound would disappear—not because land holding stops being a factor in how well women respond to the opportunity to get agricultural credits, but because women who have varying amounts of land would be equally likely to be in the treatment group or in the control group. Any bias that the amount of land causes in interpreting the results of the experiment would be distributed randomly and would be equally distributed across the groups.

But people come packaged in villages, and you can't give just some women in a small village instruction about applying for credit and not give it to others. So evaluation of these kinds of interventions are usually quasi-experiments because they have to be.

The Posttest-Only Design with Random Assignment

Look carefully at figure 5.1d. It is the second half of the Solomon four-group design and is called the **Campbell and Stanley posttest-only design**.

	<i>Time 1</i>		<i>Time 2</i>	
	<i>Assignment</i>	<i>Pretest</i>	<i>Intervention</i>	<i>Posttest</i>
Group 1	<i>R</i>		<i>X</i>	<i>O</i> ₁
Group 2	<i>R</i>			<i>O</i> ₂

Figure 5.1d. The Campbell and Stanley posttest-only design.

This design has a lot going for it. It retains the random assignment of participants in the classical design and in the Solomon four-group design, but it *eliminates pretesting*—and the possibility of a confound from pretest sensitization. When participants are assigned randomly to experimental conditions (control or treatment group), a significant difference between *O*₁ and *O*₂ in the posttest-only design means that we can have a lot of confidence that the intervention, *X*, caused that difference (Cook and Campbell 1979).

Another advantage is the huge saving in time and money. There are no pretests in this design and there are only two posttests instead of the four in the Solomon four-group design.

Here's an example of this elegant design. McDonald and Bridge (1991) asked 160 female nurses to read an information packet about a surgery patient whom they would be attending within the next 8 hours. The nurses were assigned randomly to one of eight experimental conditions: (1) The patient was named Mary B. or Robert B. This produced *two patient-gender condi-*

tions. (2) Half the nurses read only a synopsis of the condition of Mary B. or Robert B., and half read the same synopsis as the fourth one in a series of seven. This produced *two memory-load conditions*. (3) Finally, half the nurses read that the temperature of Mary B. or Robert B. had just spiked unexpectedly to 102°, and half did not. This produced *two patient stability conditions*. The three binary conditions combined to form eight experimental conditions in a **factorial design** (more on factorial designs at the end of this chapter).

Next, McDonald and Bridge asked nurses to estimate, to the nearest minute, how much time they would plan for each of several important nursing actions. Irrespective of the memory load, nurses planned significantly more time for giving the patient analgesics, for helping the patient to walk around, and for giving the patient emotional support when the patient was a man.

The posttest-only design, with random assignment, is not used as much as I think it should be, despite its elegance and its low cost. This is due partly to the appealing-but-mistaken idea that matching participants in experiments on key independent variables (age, ethnicity, etc.) is somehow better than randomly assigning participants to groups. It's also due partly to the nagging suspicion that pretests are essential.

The One-Shot Case Study

The **one-shot case study design** is shown in figure 5.1e. It is also called the **ex post facto design** because a single group of people is measured on

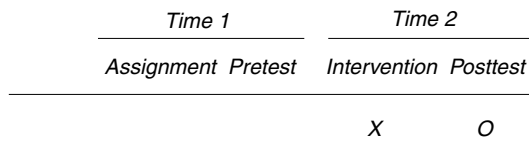


Figure 5.1e. The one-shot case study design.

some dependent variable *after* an intervention has taken place. This is the most common research design in culture change studies, where it is obviously impossible to manipulate the dependent variable. You arrive in a community and notice that something important has taken place. A clinic or a school has been built. You try to evaluate the experiment by interviewing people (*O*) and assessing the impact of the intervention (*X*).

With neither a pretest nor a control group, you can't be sure that what you observe is the result of some particular intervention. Despite this apparent weakness, however, the intuitive appeal of findings produced by one-shot case studies can be formidable.

In the 1950s, physicians began general use of the Pap Test, a simple office procedure for determining the presence of cervical cancer. Figure 5.2 shows that since 1950, the death rate from cervical cancer in the United States has

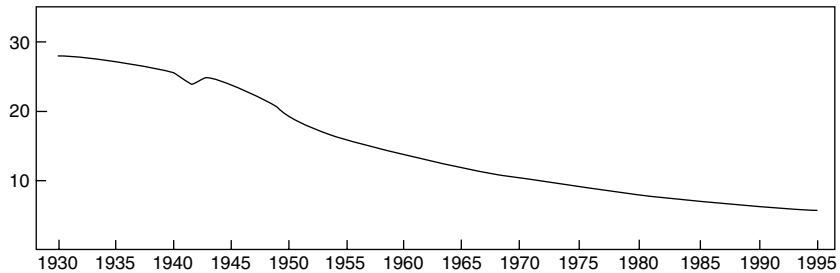


Figure 5.2. Death rate from cervical cancer, 1930–1995.

SOURCE: Adapted from B. Williams, *A Sampler on Sampling*, figure 2.1, p. 17. Reprinted with permission of Lucent Technologies Inc./Bell Labs.

dropped steadily, from about 18 per 100,000 women to about 11 in 1970, to about 8.3 in 1980, to about 6.5 in 1995 and to about 4.6 in 2000. If you look only at the data *after* the intervention (the one-shot case study X O design), you could easily conclude that the intervention (the Pap Test) was the sole cause of this drop in cervical cancer deaths. There is no doubt that the continued decline of cervical cancer deaths is due largely to the early detection provided by the Pap Test, but by 1950, the death rate had already declined by 36% from 28 per 100,000 in 1930 (Williams 1978:16).

Never use a design of less logical power when one of greater power is feasible. If pretest data are available, use them. On the other hand, a one-shot case study is often the best you can do. Virtually all ethnography falls in this category, and, as I have said before, nothing beats a good story, well told.

The One-Group Pretest-Posttest

The **one-group pretest-posttest** design is shown in figure 5.1f. Some variables are measured (observed), then the intervention takes place, and then the variables are measured again. This takes care of some of the problems associated with the one-shot case study, but it doesn't eliminate the threats of history, testing, maturation, selection, and mortality. Most importantly, if there is a significant difference in the pretest and posttest measurements, we can't tell if the intervention made that difference happen.

The one-group pretest-posttest design is commonly used in evaluating training programs. The question asked is: Did the people who were exposed to this

<i>Time 1</i>		<i>Time 2</i>	
<i>Assignment</i>	<i>Pretest</i>	<i>Intervention</i>	<i>Posttest</i>
	O_1	X	O_2

Figure 5.1f. The one-group pretest-posttest design.

skill-building program (midwives, coal miners, kindergarten teachers, etc.) get any effect from the program?

Peterson and Johnstone (1995) studied the effects on 43 women inmates of a U.S. federal prison of an education program about drug abuse. The participants all had a history of drug abuse, so there is no random assignment here. Peterson and Johnstone measured the participants' health status and perceived well-being before the program began and after the program had been running for nine months. They found that physical fitness measures were improved for the participants as were self-esteem, health awareness, and health-promoting attitudes.

The Two-Group Posttest-Only Design without Random Assignment

The **two-group posttest-only design without random assignment** design is shown in figure 5.1g. This design, also known as the **static group comparison**,

<i>Time 1</i>		<i>Time 2</i>	
<i>Assignment</i>	<i>Pretest</i>	<i>Intervention</i>	<i>Posttest</i>
		X	O_1
			O_2

Figure 5.1g. Two-group posttest-only design without random assignment: Static group comparison.

son, improves on the one-shot *ex post facto* design by adding an untreated control group—an independent case that is evaluated only at time 2. The relation between smoking cigarettes (the intervention) and getting lung cancer (the dependent variable) is easily seen by applying the humble *ex post facto* design with a control group for a second posttest.

In 1965, when the American Cancer Society did its first big Cancer Preven-

tion Study, men who smoked (that is, those who were subject to the intervention) were about 12 times more likely than nonsmokers (the control group) to die of lung cancer. At that time, relatively few women smoked, and those who did had not been smoking very long. Their risk was just 2.7 times that for women nonsmokers of dying from lung cancer.

By 1988, things had changed dramatically. Male smokers were then about 23 times more likely than nonsmokers to die of lung cancer, and female smokers were 12.8 times more likely than female nonsmokers to die of lung cancer. Men's risk had doubled (from about 12 to about 23), but women's risk had more than quadrupled (from 2.7 to about 13) (National Cancer Institute 1997). In the last decade, the death rate for lung cancer has continued to fall among men in the United States, while the death rate for women has remained about the same (<http://www.cdc.gov/cancer/lung/statistics.htm#statistics>) or even risen (Patel et al. 2004).

In true experiments run with the posttest-only design, participants are assigned at random to either the intervention or the control group. In the static-group comparison design, the researcher has no control over assignment of participants. This leaves the static-group comparison design open to an unresolvable validity threat. There is no way to tell whether the two groups were comparable at time 1, before the intervention, even with a comparison of observations 1 and 3. Therefore, you can only guess whether the intervention caused any differences in the groups at time 2.

Despite this, the static-group comparison design is the best one for evaluating natural experiments, where you have no control over the assignment of participants anyway.

Lambros Comitas and I wanted to find out if the experience abroad of Greek labor migrants had any influence on men's and women's attitudes toward gender roles when they returned to Greece. The best design would have been to survey a group before they went abroad, then again while they were away, and again when they returned to Greece. Since this was not possible, we studied one group of persons who had been abroad and another group of persons who had never left Greece. We treated these two groups as if they were part of a static-group comparison design (Bernard and Comitas 1978).

From a series of life histories with migrants and nonmigrants, we learned that the custom of giving dowry was under severe stress (Bernard and Ashton-Vouyoucalos 1976). Our survey confirmed this: Those who had worked abroad were far less enthusiastic about providing expensive dowries for their daughters than were those who had never left Greece. We concluded that this was in some measure due to the experiences of migrants in West Germany.

There were threats to the validity of this conclusion: Perhaps migrants were a self-selected bunch of people who held the dowry and other traditional

Greek customs in low esteem to begin with. But we had those life histories to back up our conclusion. Surveys are weak compared to true experiments, but their power is improved if they are conceptualized in terms of testing natural experiments and if their results are backed up with data from open-ended interviews.

Interrupted Time Series Design

The interrupted time series design, shown in figure 5.1h, can be very persuasive. It involves making a series of observations before and after an inter-

<i>Time 1</i>	<i>Time 2</i>	
<i>Assignment Pretest</i>	<i>Intervention</i>	<i>Posttest</i>
O O O	X	O O O

Figure 5.1h. The interrupted time series design.

vention. Gun-control laws are often evaluated with time series data. Loftin et al. (1991) found that the homicide rate in Washington, D.C., dropped after the D.C. City Council passed a law in 1976 restricting the sale of handguns. But it's difficult to prove cause and effect with time series data because so many other things are going on. The homicide rate dropped 15% in D.C. in 1977, the year after the law was passed, but it went down 31% in Baltimore, a nearby city of more or less the same size without similar restrictions on handgun sales (Britt et al. 1996).

Still, a well-executed time series study packs a lot of punch. On November 3, 1991, Earvin "Magic" Johnson, a star basketball player for the Los Angeles Lakers, held a news conference and announced that he was HIV-positive. By then, AIDS had been in the news for 10 years. There had been massive government and media effort to educate people about AIDS, but poll after poll showed that most people believed that AIDS was a "gay disease."

Philip Pollock (1994) treated those polls as a time series and Johnson's announcement as an interruption in the time series. After Magic Johnson's announcement, the polls began to show a dramatic change in the way people across the United States talked about AIDS (ibid.:444). If it had happened to Magic, it could have happened to anyone.

Thought Experiments

As you can see, it is next to impossible to eliminate threats to validity in natural experiments. However, there is a way to understand those threats and

to keep them as low as possible: Think about research questions as if it were possible to test them in *true* experiments. These are called **thought experiments**.

This wonderful device is part of the everyday culture in the physical sciences. In 1972, I did an ethnographic study of scientists at Scripps Institution of Oceanography. Here's a snippet from a conversation I heard among some physicists there. "If we could only get rid of clouds, we could capture more of the sun's energy to run stuff on Earth," one person said. "Well," said another, "there are no clouds above the Earth's atmosphere. The sun's energy would be lots easier to capture out there."

"Yeah," said the first, "so suppose we send up a satellite, with solar panels to convert sunlight to electricity, and we attach a really long extension cord so the satellite was tethered to the Earth. Would that work?" The discussion got weirder from there (if you can imagine), but it led to a lot of really useful ideas for research.

Suppose you wanted to know if small farms can produce organically grown food on a scale sufficiently large to be profitable. What would a true experiment to test this question look like? You might select some smallish farms with similar acreage and assign half of them randomly to grow vegetables organically. You'd assign the other half of the farms to grow the same vegetables using all the usual technology (pesticides, fungicides, chemical fertilizers, and so on). Then, after a while, you'd measure some things about the farms' productivity and profitability and see which of them did better.

How could you be sure that organic or nonorganic methods of farming made the difference in profitability? Perhaps you'd need to control for access to the kinds of market populations that are friendly toward organically produced food (like university towns) or for differences in the characteristics of soils and weather patterns. Obviously, you can't do a true experiment on this topic, randomly assigning farmers to use organic or high-tech methods, but you *can* evaluate the experiments that real farmers are conducting every day in their choice of farming practices.

So, after you've itemized the possible threats to validity in your thought experiment, go out and look for natural experiments—societies, voluntary associations, organizations—that conform most closely to your ideal experiment. Then evaluate those natural experiments.

That's what Karen Davis and Susan Weller (1999) did in their study of the efficacy of condoms in preventing the transmission of HIV among heterosexuals. Here's the experiment you'd have to conduct. First, get a thousand heterosexual couples. Make each couple randomly serodiscordant. That is, for each couple, randomly assign the man or the woman to be HIV-positive. Assign each couple randomly to one of three conditions: (1) They use condoms for

each sexual act; (2) They sometimes use condoms; or (3) They don't use condoms at all. Let the experiment run a few years. Then see how many of the couples in which condoms are always used remain serodiscordant and how many become seroconcordant—that is, they are both HIV-positive. Compare across conditions and see how much difference it makes to always use a condom.

Clearly, no one could conduct this experiment. But Davis and Weller scoured the literature on condom efficacy and found 25 studies that met three criteria: (1) The focus was on serodiscordant heterosexual couples who said they regularly had penetrative sexual intercourse; (2) The HIV status of the participants in each study had been determined by a blood test; and (3) There was information on the use of condoms. The 25 studies involved thousands of participants, and from this **meta-analysis** Davis and Weller established that consistent use of condoms reduced the rate of HIV transmission by over 85%.

If you look around, there are natural experiments going on all around you. On January 1, 2002, 12 of the 15 members of the European Union gave up their individual currencies and adopted the euro. Greece was one of the 12, Denmark wasn't. Some researchers noticed that many of the euro coins were smaller than the Greek drachma coins they'd replaced and thought that this might create a choking hazard for small children (Papadopoulos et al. 2004). The researchers compared the number of choking incidents reported in Danish and Greek hospitals in January through March from 1996 through 2002. Sure enough, there was no increase in the rate of those incidents in Denmark (which hadn't converted to the euro), but the rate in Greece suddenly more than doubled in 2002.

True Experiments in the Lab

Laboratory experiments to test theories about how things work in the real world is the preeminent method in social psychology. Cognitive dissonance theory, for example, predicts that people who come out of a tough initiation experience (marine recruits at boot camp, prisoners of war, girls and boys who go through genital mutilation, etc.) wind up as supporters of their tormentors. In a classic experiment, Elliot Aronson and Judson Mills (1959) recruited 63 college women for a discussion group that ostensibly was being formed to talk about psychological aspects of sex. To make sure that only mature people—people who could discuss sex openly—would make it into this group, some of the women would have to go through a screening test. Well, that's what they were *told*.

A third of the women were assigned randomly to a group that had to read a

list of obscene words and some sexually explicit passages from some novels—aloud, in front of a man who was running the experiment. (It may be hard to imagine now, but those women who went through this in the 1950s must have been very uncomfortable.) Another third were assigned randomly to a group that had to recite some nonobscene words that had to do with sex and a third group went through no screening at all.

Then, each participant was given headphones to listen in on a discussion that was supposedly going on among the members of the group she was joining. The “discussion” was actually a tape and it was, as Aronson and Mills said, “one of the most worthless and uninteresting discussions imaginable” (ibid.:179). The women rated the discussion, on a scale of 0–15, on things like dull-interesting, intelligent-unintelligent, and so on.

Those in the tough initiation condition rated the discussion higher than did the women in either the control group or the mild initiation group. Since all the women were assigned randomly to participate in one of the groups, the outcome was unlikely to have occurred by chance. Well, the women in the tough initiation condition had gone through a lot to join the discussion. When they discovered how boringly nonprurient it was, what did they do? They convinced themselves that the group was worth joining.

Aronson and Mills’s findings were corroborated by Gerard and Mathewson (1966) in an independent experiment. Those findings from the laboratory can now be the basis for a field test, across cultures, of the original hypothesis.

Conversely, events in the real world can stimulate laboratory experiments. In 1963, in Queens, New York, Kitty Genovese was stabbed to death in the street one night. There were 38 eyewitnesses who saw the whole grisly episode from their apartment windows, and not one of them called the police. The newspapers called it “apathy,” but Bibb Latané and John Darley had a different explanation. They called it **diffusion of responsibility** and they did an experiment to test their idea (1968).

Latané and Darley invited ordinary people to participate in a “psychology experiment.” While the participants were waiting in an anteroom to be called for the experiment, the room filled with smoke. If there was a single participant in the room, 75% reported the smoke right away. If there were three or more participants waiting together, they reported the smoke only 38% of the time. People in groups just couldn’t figure out whose responsibility it was to do something. So they did nothing.

True Experiments in the Field

When experiments are done outside the lab, they are called **field experiments**. Jacob Hornik (1992) does experiments to test the effect of being

touched on consumers' spending. In the first study, as lone shoppers (no couples) entered a large bookstore, an "employee" came up and handed them a catalog. Alternating between customers, the employee-experimenter touched about half the shoppers lightly on the upper arm. The results? Across 286 shoppers, those who were touched spent an average of \$15.03; those who were not touched spent just \$12.23. And this difference was across the board, no matter what the sex of the toucher or the shopper.

In another of his experiments, Hornik enlisted the help of eight servers—four waiters and four waitresses—at a large restaurant. At the end of the meal, the servers asked each of 248 couples (men and women) how the meal was. Right then, for half the couples, the servers touched the arm of either the male or the female in the couple for one second. The servers didn't know it, but they had been selected out of 27 waiters and waitresses in the restaurant to represent two ends of a physical attractiveness scale. The results? Men and women alike left bigger tips when they were touched, but the effect was stronger for women patrons than it was for men. Overall, couples left about a 19% tip when the woman was touched, but only 16.5% when the man was touched.

Field experiments like these can produce powerful evidence for applications projects, but to be really useful, you need to back up results with ethnography so that you understand the process. How did all this tipping behavior play out? Did women who were touched suggest a bigger tip when the time came to pay the bill, or did the men suggest the tip? Or did it all happen without any discussion? Rich ethnographic data are needed here.

Field experiments are rare in anthropology, but not unprecedented. Marvin Harris and his colleagues (1993) ran an experiment in Brazil to test the effect of substituting one word in the question that deals with race on the Brazilian census. The demographers who designed the census had decided that the term *parda* was a more reliable gloss than the term *morena* for what English speakers call "brown," despite overwhelming evidence that Brazilians prefer the term *morena*.

In the town of Rio de Contas, Harris et al. assigned 505 houses randomly to one of two groups and interviewed one adult in each house. All respondents were asked to say what *cor* (color) they thought they were. This was the "free-choice option." Then they were asked to choose one of four terms that best described their *cor*. One group (with 252 respondents) was asked to select among *branca* (white), *parda* (brown), *preta* (black), and *amerela* (yellow). This was the "*parda* option"—the one used on the Brazilian census. The other group (with 253 respondents) was asked to select among *branca*, *morena* (brown), *preta*, and *amerela*. This was the "*morena* option," and is the intervention, or treatment in Harris's experiment.

Among the 252 people given the *parda* option, 131 (52%) identified themselves as *morena* in the free-choice option (when simply asked to say what color they were). But when given the *parda* option, only 80 of those people said they were *parda* and 41 said they were *branca* (the rest chose the other two categories). Presumably, those 41 people would have labeled themselves *morena* if they'd had the chance; not wanting to be labeled *parda*, they said they were *branca*. The *parda* option, then, produces more Whites (*brancas*) in the Brazilian census and fewer browns (*pardas*).

Of the 253 people who responded to the *morena* option, 160 (63%) said they were *morena*. Of those 160, only 122 had chosen to call themselves *morena* in the free-choice option. So, giving people the *morena* option actually increases the number of browns (*morenas*) and decreases the number of Whites (*brancas*) in the Brazilian census.

Does this difference make a difference? Demographers who study the Brazilian census have found that those who are labeled Whites live about 7 years longer than do those labeled non-Whites in that country. If 31% of self-described *morenas* say they are Whites when there is no *morena* label on a survey and are forced to label themselves *parda*, what does this do to all the social and economic statistics about racial groups in Brazil? (Harris et al. 1993).

Natural Experiments

True experiments and quasi-experiments are *conducted* and the results are *evaluated* later. **Natural experiments**, by contrast, are going on around us all the time. They are not conducted by researchers at all—they are simply evaluated.

Here are four examples of common natural experiments: (1) Some people choose to migrate from villages to cities, while others stay put. (2) Some villages in a region are provided with electricity, while some are not. (3) Some middle-class Chicano students go to college, some do not. (4) Some cultures practice female infanticide, some do not.

Each of these situations is a natural experiment that tests *something* about human behavior and thought. The trick is to ask: “What hypothesis is being tested by what’s going on here?”

To evaluate natural experiments—that is, to figure out what hypothesis is being tested—you need to be alert to the possibilities and collect the right data. There’s a really important natural experiment going in an area of Mexico where I’ve worked over the years. A major irrigation system has been installed over the last 40 years in parts of the Mezquital, a high desert valley. Some of

the villages affected by the irrigation system are populated entirely by Ñähñu (Otomí) Indians; other villages are entirely mestizo (as the majority population of Mexico is called).

Some of the Indian villages in the area are too high up the valley slope for the irrigation system to reach. I could not have decided to run this multimillion dollar system through certain villages and bypass others, but the instant the decision was made by others, a natural experiment on the effects of a particular intervention was set in motion. There is a treatment (irrigation), there are treatment groups (villages full of people who get the irrigation), and there are control groups (villages full of people who are left out).

Unfortunately, I can't evaluate the experiment because I simply failed to see the possibilities early enough. Finkler (1974) saw the possibilities; her ethnographic study of the effects of irrigation on an Indian village in the same area shows that the intervention is having profound effects. But neither she nor I measured (pretested) things like average village wealth, average personal wealth, migration rates, alcoholism, etc., that I believe have been affected by the coming of irrigation.

Had anyone done so—if we had **baseline data**—we would be in a better position to ask: “What hypotheses about human behavior are being tested by this experiment?” I can't reconstruct variables from 20 or 30 years ago. The logical power of the experimental model for establishing cause and effect between the intervention and the dependent variables is destroyed.

Some natural experiments, though, produce terrific data all by themselves for evaluation. In 1955, the governor of Connecticut ordered strict enforcement of speeding laws in the state. Anyone caught speeding had his or her driver's license suspended for at least 30 days. Traffic deaths fell from 324 in 1955 to 284 in 1956. A lot of people had been inconvenienced with speeding tickets and suspension of driving privileges, but 40 lives had been saved.

Did the crackdown cause the decline in traffic deaths? Campbell and Ross (1968) used the available data to find out. They plotted the traffic deaths from 1951 to 1959 in Connecticut, Massachusetts, New York, New Jersey, and Rhode Island. Four of the five states showed an increase in highway deaths in 1955, and all five states showed a decline in traffic deaths the following year, 1956. If that were all you knew, you couldn't be sure about the cause of the decline. However, traffic deaths continued to decline steadily in Connecticut for the next 3 years (1957, 1958, 1959). In Rhode Island and Massachusetts they went up; in New Jersey, they went down a bit and then up again; and in New York, they remained about the same.

Connecticut was the only state that showed a consistent reduction in highway deaths for 4 years after the stiff penalties were introduced. Campbell and

Ross treated these data as a series of natural experiments, and the results were convincing: Stiff penalties for speeders saves lives.

Natural Experiments Are Everywhere

If you think like an experimentalist, you eventually come to see the unlimited possibilities for research going on all around you. For example, Cialdini et al. (1976) evaluated the natural experiment in pride that is conducted on most big university campuses every weekend during football season. Over a period of 8 weeks, professors at Arizona State, Louisiana State, Ohio State, Notre Dame, Michigan, the University of Pittsburgh, and the University of Southern California recorded the percentage of students in their introductory psychology classes who wore school insignias (buttons, hats, t-shirts, etc.) on the Monday after Saturday football games. For 177 students per week, on average, over 8 weeks, 63% wore some school insignia after wins in football vs. 44% after losses or ties. The difference is statistically significant.

And Kathy Oths (1994) turned what could have been a killer history conundrum into an evaluation of a natural experiment in her work on medical choice in Peru. From July 1 to December 15, 1988, Oths visited each of the 166 households in Chugurpampa, Peru, several times to collect illness histories—what kinds of illnesses people came down with and what they did about them. When she began the project, the harvest season had just ended in the high Andes and the farmers in Chugurpampa had collected most of the money they would have for the year. But in August, a month into her work, catastrophic inflation hit Peru and the local currency, the Peruvian inti, which was at 40 to the U.S. dollar, was crushed. It hit 683 by November.

Oths continued her household visits. As the hard times dragged on, the people of Chugurpampa continued to report the same number of illnesses (between seven and eight per household per month) but they defined a larger percentage of their illnesses as mild (requiring only home remedies) and a smaller percentage as moderate or severe (requiring visits to doctors or to the hospital). In other words, they spent what little money they had on cases that they thought needed biomedical intervention and stopped spending money on traditional healers.

Naturalistic Experiments

In a **naturalistic** experiment, you *contrive* to collect experimental data under natural conditions. You make the data happen, out in the natural world (not in the lab), and you evaluate the results.

In a memorable experiment, elegant in its simplicity of design, Doob and Gross (1968) had a car stop at a red light and wait for 15 seconds after the light turned green before moving again. In one experimental condition, they used a new car and a well-dressed driver. In another condition, they used an old, beat-up car and a shabbily dressed driver. They repeated the experiment many times and measured the time it took for people in the car behind the experimental car to start honking their horns. It won't surprise you to learn that people were quicker to vent their frustration at apparently low-status cars and drivers.

Piliavin et al. (1969) did a famous naturalistic experiment to test the "good Samaritan" problem. Students in New York City rode a particular subway train that had a 7.5-minute run at one point. At 70 seconds into the run, a researcher pitched forward and collapsed. The team used four experimental conditions: The "stricken" person was either black or white and was either carrying a cane or a liquor bottle. Observers noted how long it took for people in the subway car to come to the aid of the supposedly stricken person, the total population of the car, whether bystanders were black or white, and so on. You can conjure up the results. There were no surprises.

Harari et al. (1985) recruited drama majors to test whether men on a college campus would come to the aid of a woman being raped. They staged realistic-sounding rape scenes and found that there was a significant difference in the helping reaction of male passersby if those men were alone or in groups.

The Small-World Experiment

Consider this: You're having coffee near the Trevi Fountain in Rome. You overhear two Americans chatting next to you and you ask where they're from. One of them says she's from Sioux City, Iowa. You say you've got a friend from Sioux City and it turns out to be your new acquaintance's cousin. The culturally appropriate reaction at this point is for everyone to say, "Wow, what a small world!"

Stanley Milgram (1967) invented an experiment to test how small the world really is. He asked a group of people in the midwestern United States to send a folder to a divinity student at Harvard University, but only if the participant *knew* the divinity student personally. Otherwise, Milgram asked people to send the folders to an acquaintance whom they thought had a chance of knowing the "target" at Harvard.

The folders got sent around from acquaintance to acquaintance until they wound up in the hands of someone who actually knew the target—at which point the folders were sent, as per the instructions in the game, to the target.

The average number of *links* between all the “starters” and the target was about five. It really *is* a small world.

No one expects this experiment to actually happen in real life. It’s contrived as can be and it lacks control. But it’s compelling because it says *something* about how the natural world works. The finding was so compelling that it was the basis for the Broadway play “Six Degrees of Separation,” as well as the movie of the same name that followed and the game “Six Degrees of Kevin Bacon.” Tell people who don’t know about the six-degrees-of-separation phenomenon about Milgram’s experiment and ask them to guess how many links it takes to get a folder between any two randomly chosen people in the United States. Most people will guess a much bigger number than five.

The Lost-Letter Technique

Another of Milgram’s contributions is a method for doing unobtrusive surveys of political opinion. The method is called the “lost-letter technique” and consists of “losing” a lot of letters that have addresses and stamps on them (Milgram et al. 1965).

The technique is based on two assumptions. First, people in many societies believe that they ought to mail a letter if they find one, especially if it has a stamp on it. Second, people will be less likely to drop a lost letter in the mail if it is addressed to someone or some organization that they don’t like.

Milgram et al. (ibid.) tested this in New Haven, Connecticut. They lost 400 letters in 10 districts of the city. They dropped the letters on the street; they left them in phone booths; they left them on counters at shops; and they tucked them under windshield wipers (after penciling “found near car” on the back of the envelope). Over 70% of the letters addressed to an individual or to a medical research company were returned. Only 25% of the letters addressed to either “Friends of the Communist Party” or “Friends of the Nazi Party” were returned. (The addresses were all the same post box that had been rented for the experiment.)

By losing letters in a sample of communities, then, and by counting the differential rates at which they are returned, you can test variations in sentiment. Two of Milgram’s students distributed anti-Nazi letters in Munich. The letters did not come back as much from some neighborhoods as from others, and they were thus able to pinpoint the areas of strongest neo-Nazi sentiment (Milgram 1969:68). The lost-letter technique has sampling problems and validity problems galore associated with it. But it’s still an interesting way to infer public opinion about emotionally charged issues, and you can see just how intuitively powerful the results can be. (For more examples of the lost-letter technique, see Hedge and Yousif 1992 and Bridges and Coady 1996.)

Comparative Field Experiments

Naturalistic field experiments appeal to me because they are excellent for comparative research, and comparison is so important for developing theory. Feldman (1968) did five field experiments in Paris, Boston, and Athens to test whether people in those cities respond more kindly to foreigners or to members of their own culture.

In one experiment, the researchers simply asked for directions and measured whether foreigners or natives got better treatment. Parisians and Athenians gave help significantly more often to fellow citizens than to foreigners. In Boston, there was no difference.

In the second experiment, foreigners and natives stood at major metro stops and asked total strangers to do them a favor. They explained that they were waiting for a friend, couldn't leave the spot they were on, and had to mail a letter. They asked people to mail the letters for them (the letters were addressed to the experiment headquarters) and simply counted how many letters they got back from the different metro stops in each city. Half the letters were unstamped.

In Boston and Paris, between 32% and 35% of the people refused to mail a letter for a fellow citizen. In Athens, 93% refused. Parisians treated Americans significantly better than Bostonians treated Frenchmen on this task. In fact, in cases where Parisians were asked to mail a letter that was stamped, they treated Americans significantly better than they treated other Parisians! (So much for *that* stereotype.)

In the third experiment, researchers approached informants and said: "Excuse me, sir. Did you just drop this dollar bill?" (or other currency, depending on the city). It was easy to measure whether or not people falsely claimed the money more from foreigners than from natives. This experiment yielded meager results.

In the fourth experiment, foreigners and natives went to pastry shops in the three cities, bought a small item and gave the clerk 25% more than the item cost. Then they left the shop and recorded whether the clerk had offered to return the overpayment. This experiment also showed little difference among the cities, or between the way foreigners and locals are treated.

And in the fifth experiment, researchers took taxis from the same beginning points to the same destinations in all three cities. They measured whether foreigners or natives were charged more. In neither Boston nor Athens was a foreigner overcharged more than a local. In Paris, however, Feldman found that "the American foreigner was overcharged significantly more often than the French compatriot in a variety of ingenious ways" (1968:11).

Feldman collected data on more than 3,000 interactions and was able to

draw conclusions about cultural differences in how various peoples respond to foreigners as opposed to other natives. Some stereotypes were confirmed, while others were crushed.

Bochner did a series of interesting experiments on the nature of Aboriginal-white relations in urban Australia (see Bochner [1980:335–340] for a review). These experiments are clever, inexpensive, and illuminating, and Bochner's self-conscious critique of the limitations of his own work is a model for field experimentalists to follow. In one experiment, Bochner put two classified ads in a Sydney paper:

Young couple, no children, want to rent small unfurnished flat up to \$25 per week. Saturday only. 759–6000.

Young Aboriginal couple, no children, want to rent small unfurnished flat up to \$25 per week. Saturday only. 759–6161. (Bochner 1972:335)

Different people were assigned to answer the two phones, to ensure that callers who responded to both ads would not hear the same voice. Note that the ads were identical in every respect, except for fact that in one of the ads the ethnicity of the couple was identified, while in the other it was not. There were 14 responses to the ethnically nonspecific ad and two responses to the ethnically specific ad (three additional people responded to both ads).

In another experiment, Bochner exploited what he calls the "Fifi effect" (Bochner 1980:336). The Fifi effect refers to the fact that urbanites acknowledge the presence of strangers who pass by while walking a dog and ignore others. Bochner sent a white woman and an Aboriginal woman, both in their early 20s, and similarly dressed, to a public park in Sydney. He had them walk a small dog through randomly assigned sectors of the park, for 10 minutes in each sector.

Each woman was followed by two observers, who gave the impression that they were just out for a stroll. The two observers *independently* recorded the interaction of the women with passersby. The observers recorded the frequency of smiles offered to the women; the number of times anyone said anything to the women; and the number of nonverbal recognition nods the women received. The white woman received 50 approaches, while the Aboriginal woman received only 18 (Bochner 1971:111).

There are many elegant touches in this experiment. Note how the age and dress of the experimenters were controlled, so that only their ethnic identity remained as a dependent variable. Note how the time for each experimental trial (10 minutes in each sector) was controlled to ensure an equal opportunity for each woman to receive the same treatment by strangers. Bochner did preliminary observation in the park and divided it into sectors that had the same

population density, so that the chance for interaction with strangers would be about equal in each run of the experiment, and he used two independent observer-recorders.

As Bochner points out, however, there were still design flaws that threatened the internal validity of the experiment (1980:337). As it happens, the interrater reliability of the two observers in this experiment was nearly perfect. But suppose the two observers shared the same cultural expectations about Aboriginal-white relations in urban Australia. They might have quite reliably misrecorded the cues that they were observing.

Reactive and unobtrusive observations alike tell you *what* happened, not *why*. It is tempting to conclude that the Aboriginal woman was ignored because of active prejudice. But, says Bochner, “perhaps passersby ignored the Aboriginal . . . because they felt a personal approach might be misconstrued as patronizing” (ibid.:338).

In Bochner’s third study, a young white or Aboriginal woman walked into a butcher’s shop and asked for 10 cents’ worth of bones for her pet dog. The dependent variables in the experiment were the weight and quality of the bones. (An independent dog fancier rated the bones on a 3-point scale, without knowing how the bones were obtained, or *why*.) Each woman visited seven shops in a single middle-class shopping district.

In both amount and quality of bones received, the white woman did better than the Aboriginal, but the differences were not statistically significant—the sample was just too small so no conclusions could be drawn from that study alone. *Taken all together*, though, the three studies done by Bochner and his students comprise a powerful set of information about Aboriginal-white relations in Sydney. Naturalistic experiments have their limitations, but they often produce intuitively compelling results.

Are Field Experiments Ethical?

Field experiments come in a range of ethical varieties, from innocuous to borderline to downright ugly. I see no ethical problems with the lost-letter technique. When people mail one of the lost letters, they don’t know that they are taking part in an experiment, but that doesn’t bother me. Personally, I see no harm in the experiment to test whether people vent their anger by honking their car horns more quickly at people they think are lower socioeconomic class. These days, however, with road rage an increasing problem, I don’t recommend repeating Doob and Gross’s experiment.

Randomized field experiments, used mostly in evaluation research, can be problematic. Suppose you wanted to know whether fines or jail sentences are

better at changing the behavior of drunk drivers. One way to do that would be to randomly assign people who were convicted of the offense to one or the other condition and watch the results. Suppose one of the participants whom you didn't put in jail kills an innocent person?

The classic experimental design in drug testing requires that some people get the new drug, that some people get a placebo (a sugar pill that has no effect), and that neither the patients nor the doctors administering the drugs know which is which. This double-blind placebo design is responsible for great advances in medicine and the saving of many lives. But suppose that, in the middle of a double-blind trial of a drug you find out that the drug really works. Do you press on and complete the study? Or do you stop right there and make sure that you aren't withholding treatment from people whose lives could be saved? The ethical problems associated with **withholding of treatment** are under increasing scrutiny (see, for example, Wertz 1987; De Leon et al. 1995; Miller and Shorr 2002; and Storosum et al. 2003).

There is a long history of debate about the ethics of **deception** in psychology and social psychology (see Korn [1997] for a review). My own view is that, on balance, some deception is clearly necessary—certain types of research just can't be done without it. When you use deception, though, you run all kinds of risks—not just to research participants, but to the research itself. These days, college students (who are the participants for most social psych experiments) are very savvy about all this and are on the lookout for clues as to the “real” reason for an experiment the minute they walk in the door.

If you don't absolutely need deception in true behavioral experiments, that's one less problem you have to deal with. If you decide that deception is required, then understand that the responsibility for any bad outcomes is yours and yours alone.

The experiments by Piliavin et al. (1969) and Harari et al. (1985) on whether people will come to the aid of a stricken person, or a woman being raped, present real ethical problems. Some of the participants (who neither volunteered to be in an experiment nor were paid for their services) might still be wondering what happened to that poor guy on the subway whom they stepped over in their hurry to get away from an uncomfortable situation—or that woman whose screams they ignored. In laboratory experiments, at least, participants are debriefed—told what the real purpose of the study was—in order to reduce emotional distress. In the guerrilla theater type of field experiment, though, no debriefing is possible.

Even debriefing has its dark side. People don't like to find out that they have been duped into being part of an experiment, and some people may suffer a terrible loss of self-esteem if they do find out and conclude that they acted

badly. How would *you* feel if you were one of the people who failed to respond to a rape victim and then were told that you were just part of an experiment—that no real rape ever took place, and thank you very much for your help?

If you think some of *these* cases are borderline, consider the study by West et al. (1975) on whether there is a little larceny in us all.

The Watergate Experiment

In the Watergate affair, men loyal to then President Richard Nixon broke into the headquarters of the Democratic Party at the Watergate Hotel in Washington, D.C., to photograph documents pertinent to the 1972 election campaign. Their bungling of the job, and the subsequent cover-up by Nixon and his staff at the White House, led to the unprecedented resignation of the president of the United States from office in 1974. Soon thereafter, West et al. conducted their experiment.

They confronted 80 different students with a proposition to burglarize a local advertising firm. Participants were randomly assigned to one of four conditions. In the first condition, participants were told that the job was to be done for the Internal Revenue Service. The IRS, it seemed, needed to get the goods on this company in order to bring them to trial for tax evasion. If the participants were caught in the act, then the government would guarantee immunity from prosecution. In the second condition, participants were told that there was no immunity from prosecution.

In the third condition, participants were told that another advertising agency had paid \$8,000 for the job, and that they (the participants) would get \$2,000 for their part in it. (Remember, that was \$2,000 in 1979—about \$8,000 today.) Finally, in the fourth condition, participants were told that the burglary was being committed just to see if the plan would work. Nothing would be taken from the office.

Understand that this was not a “let’s pretend” exercise. Participants were not brought into a laboratory and told to imagine that they were being asked to commit a crime. This was for real. Participants met the experimenter at his home or at a restaurant. They were all criminology students at a university and knew the experimenter to be an actual local private investigator. The private eye arranged an elaborate and convincing plan for the burglary, including data on the comings and goings of police patrol cars, aerial photographs, blueprints of the building—the works.

The participants really believed that they were being solicited to commit a crime. Just as predicted by the researchers, a lot of them agreed to do it in the

first condition, when they thought the crime was for a government agency and that they'd be free of danger from prosecution if caught. What do you suppose would happen to *your* sense of self-worth when you were finally debriefed and told that you were one of the 36 out of 80 (45%) who agreed to participate in the burglary in the first condition? (See Cook [1975] for a critical comment on the ethics of this experiment.)

The key ethical issue in the conduct of all social research is whether those being studied are placed at risk by those doing the studying. This goes for field research—including surveys, ethnographies, and naturalistic experiments—as much as it does for laboratory studies. All universities in the United States have long had Institutional Review Boards, or IRBs. These are internal agencies whose members review and pass judgment on the ethical issues associated with all research on people, including biomedical and psychosocial. The concept of informed consent has developed and matured over the years. All researchers are asked by the IRBs to describe clearly and precisely what steps will be taken to ensure that people who participate in research will be protected from harm. And not just physical harm. Research participants should not experience emotional harm or financial harm, either.

Factorial Designs: Main Effects and Interaction Effects

Most experiments involve analyzing the effects of several independent variables at once. A **factorial design** lays out all the combinations of all the categories of the independent variables. That way you know how many participants you need, how many to assign to each condition, and how to run the analysis when the data are in.

It is widely believed that a good laugh has healing power. Rotton and Shats (1996) developed an experimental design to test this. They recruited 39 men and 39 women who were scheduled for orthopedic surgery. The patients were assigned randomly to one of nine groups—eight experimental groups and one control group. The patients in the eight treatment groups got to watch a movie in their room the day after their surgery.

There were three variables: choice, humor, and expectancy. The participants in the high-choice group got a list of 20 movies from which they chose four. The participants in the low-choice group watched a movie that one of the people in the high-choice group had selected. Half the participants watched humorous movies, and half watched action or adventure movies. Before watching their movie, half the participants read an article about the benefits of humor, while half read an article about the healthful benefits of exciting movies.

Figure 5.3 is a branching tree diagram that shows how these three variables, each with two attributes, create the eight logical groups for Rotton and Shats's

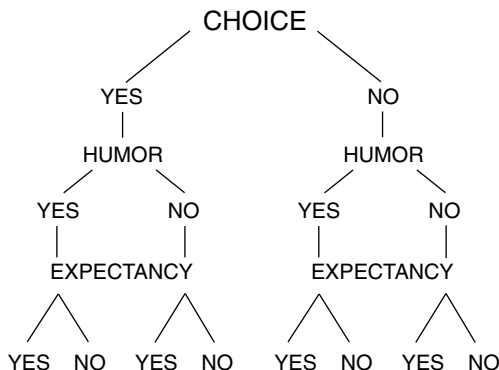


Figure 5.3. The eight conditions in Rotton and Shat's $2 \times 2 \times 2$ design.
 SOURCE: E. G. Cohn and J. Rotton, "Assault as a Function of Time and Temperature: A Moderator-Variable Time-Series Analysis," *Journal of Personality and Social Psychology*, Vol. 72, pp. 1322–1334, 1997, American Psychological Association, reprinted with permission.

experiment. Table 5.1 shows the same eight-group design, but in a format that is more common. The eight nodes at the bottom of the tree in figure 5.3 and the sets of numbers in the eight boxes of table 5.1 are called **conditions**. For example, in condition 1 (at the bottom left of figure 5.3), the patients had a choice of movie; they saw a humorous movie; and they had been led to expect that humor has healing benefits.

The dependent variables in this study included a self-report by patients on

TABLE 5.1
 Three-Way, $2 \times 2 \times 2$, Factorial Design

		Variable 3		
		Attribute 1	Attribute 2	
Variable 1	Attribute 1	Variable 2	Attribute 1	Attribute 2
		Attribute 1	1,1,1 Condition 1	1,1,2 Condition 2
	Attribute 2	1,2,1 Condition 3	1,2,2 Condition 4	
	Attribute 2	Attribute 1	2,1,1 Condition 5	2,1,2 Condition 6
Attribute 2		2,2,1 Condition 7	2,2,2 Condition 8	

the amount of pain they had and a direct measure of the amount of pain medication they took. All the patients had access to a device that let them administer more or less of the analgesics that are used for controlling pain after orthopedic surgery.

In assessing the results of a factorial experiment, researchers look for **main effects** and **interaction effects**. Main effects are the effects of each independent variable on each dependent variable. Interaction effects are effects on dependent variables that occur as a result of *interaction* between two or more independent variables. In this case, Rotton and Shats wanted to know the effects of humor on postoperative pain, but they wanted to know the effect *in different contexts*: in the context of choosing the vehicle of humor or not; in the context of being led to believe that humor has healing benefits or not; and so on.

As it turned out, being able to choose their own movie had no effect when patients saw action films. But patients who saw humorous films and who had not been able to make their own choice of film gave themselves more pain killer than did patients who saw humorous films and had been able to make the selection themselves (Rotton and Shats 1996).

We'll look at how to measure these effects when we take up ANOVA, or analysis of variance, in chapter 20.

6



Sampling

What Are Samples and Why Do We Need Them?

Informant accuracy, data validity, and ethical questions—like whether it’s alright to deceive people in conducting experiments—are all measurement problems in research. The other big class of problems involves sampling: Given that your measurements are credible, how much of the world do they represent? How far can you generalize the results of your research?

The answer depends, first of all, on the kind of data in which you’re interested. There are two kinds of data of interest to social scientists: **individual attribute data** and **cultural data**. These two kinds require different approaches to sampling.

Individual data are about *attributes of individuals in a population*. Each person has an age, for example; each person has an income; and each person has preferences for things like characteristics of a mate. If the idea is to estimate the average age, or income, or preference in a population—that is, to estimate some **population parameters**—then a scientifically drawn, unbiased sample is a must.

By “scientifically drawn,” I mean random selection of cases so that *every unit of analysis has an equal chance of being chosen for study*.

Cultural data are different. We expect cultural facts to be shared and so cultural data require experts. If you want to understand a process—like breastfeeding, or the making up of a guest list for a wedding, or a shaman’s treatment of a sick person—then you want people who can offer expert explanations about the cultural norm and about variations on that norm. It’s one thing to ask: “How many cows did *you* give to your in-laws as bride price when you got married?” It’s quite another thing to ask: “So, why do men who get mar-

ried around here deliver cows to their in-laws? . . . And how many do men usually give?”

Individual attribute data require **probability sampling**; cultural data require **nonprobability sampling**. This chapter is about probability sampling, which will take us into a discussion of probability theory, variance, and distributions. We'll get to nonprobability sampling in chapter 7.

Why the United States Still Has a Census

If samples were just easier and cheaper to study but failed to produce useful data, there wouldn't be much to say for them. A study based on a random sample, however, is often *better* than one based on the whole population.

Once every 10 years, since 1790, the United States has conducted a census in which every person in the country is supposed to be counted, in order to apportion seats in the House of Representatives to the states. Lots of things can go wrong with counting. Heads of households are responsible for filling out and returning the census forms, but in 1990, only 63% of the mailed forms were returned, and that was down from 78% in 1970. The Bureau of the Census had to hire and train half a million people to track down all the people who had not been enumerated in the mailed-back forms. Even then, there were problems with the final numbers. Some college students were counted twice: Their parents had counted them on the mailed-back census form and then, on census day, some of those same students were tracked down again by enumerators who canvassed the dorms. Meanwhile, lots of other people (like illegal immigrants and people living in places to which the census takers would rather not go) were not being counted at all.

In 1997, the Bureau of the Census asked the U.S. Congress to allow sampling instead of counting for at least some parts of the 2000 Census. This caused a serious political problem: If sampling produced more accurate (and, presumably, higher) estimates of the number of citizens who are, say, homeless or who are migrant farm workers, this would benefit only certain states. So, Congress rejected the proposal, citing the Article 1, Section 2 of the Constitution, which calls the Census an “actual Enumeration” (with a capital E, no less).

No getting around it: Actually enumerating means counting, not estimating, and the U.S. Supreme Court agreed, in 1999. To deal with the inaccuracies of a head count, the Bureau of the Census publishes adjustment tables, based on samples. In 2000, for example, the Bureau determined that it had undercounted American Indians who live off reservations by about 57,000 (see U.S. Bureau of the Census n.d.).

It Pays to Take Samples and to Stick with Them

If you are doing all the work yourself, it's next to impossible to interview more than a few hundred people. Even in a community of just 1,000 households, you'd need several interviewers to reach everyone. Interviewers may not use the same wording of questions; they may not probe equally well on subjects that require sensitive interviewing; they may not be equally careful in recording data on field instruments and in coding data for analysis. The more personnel there are on any project, the greater the instrumentation threat and the more risk to the validity of the data.

Most important, you have no idea how much error is introduced by these problems. A well-chosen sample, interviewed by people who have similarly high skills in getting data, has a known chance of being incorrect on any variable. (Careful, though: If you have a project that requires multiple interviewers and you try to skimp on personnel, you run a big risk. Overworked or poorly trained interviewers will cut corners; see chapter 8.)

Furthermore, studying an entire population may pose a history threat to the internal validity of your data. If you *don't* add interviewers it may take you so long to complete your research that events intervene that make it impossible to interpret your data.

For example, suppose you're interested in how a community of Hopi feel about certain aspects of the relocation agreement being forged in their famous land dispute with the Navajo (Brugge 1999). You decide to interview *all* 210 adults in the community. It's difficult to get some people at home, but you figure that you'll just do the survey, a little at a time, while you're doing other things during your year in the field.

About 6 months into your fieldwork, you've gotten 160 interviews on the topic—only 50 to go. Just about that time, the courts adjudicate a particularly sore point that has been in dispute for a decade regarding access to a particular sacred site. All of a sudden, the picture changes. Your "sample" of 160 is biased toward those people whom it was easy to find, and you have no idea what *that* means. And even if you could now get those remaining 50 informants, their opinions may have been radically changed by the court judgment. The opinions of the 160 informants who already talked to you may have also changed.

Now you're really stuck. You can't simply throw together the 50 and the 160, because you have no idea what that will do to your results. Nor can you compare the 160 and the 50 as representing the community's attitudes before and after the judgment. Neither sample is representative of the community.

If you had taken a representative sample of 60 people in a single week early in your fieldwork, you'd now be in much better shape, because you'd know

the potential sampling error in your study. (I'll discuss how you know this later on in this chapter.) When historical circumstances (the surprise court judgment, for example) require it, you could interview the same sample of 60 again (in what is known as a **panel study**), or take another representative sample of the same size and see what differences there are before and after the critical event.

In either case, *you are better off with the sample than with the whole population*. By the way, there is no guarantee that a week is quick enough to avoid the problem described here. It's just less likely to be a problem.

Sampling Frames

If you can get it, the first thing you need for a good sample is a good **sampling frame**. (I say, "if you can get it," because a lot of social research is done on populations for which no sampling frame exists. More on this at the end of this chapter.) A sampling frame is a list of units of analysis *from which* you take a sample and *to which* you generalize.

A sampling frame may be a telephone directory, or the tax rolls of a community, or a census of a community that you do yourself. In the United States, the city directories (published by R. L. Polk and Company) are often adequate sampling frames. The directories are available for many small towns at the local library or Chamber of Commerce. Professional survey researchers in the United States often purchase samples from firms that keep up-to-date databases just for this purpose.

For many projects, though, especially projects that involve field research, you have to get your own census of the population you are studying. Whether you work in a village or a hospital, a census gives you a sampling frame from which to take many samples during a research project. It also gives you a basis for comparison if you go back to the same community later.

Simple Random Samples

To get a **simple random sample** of 200 out of 640 people in a village, you number each individual from 1 to 640 and then take a random grab of 200 out of the numbers from 1 to 640. The easiest way to take random samples is with a computer. All of the popular program packages for statistical analysis have built-in random-number generators. Some of the most popular include SAS®, SPSS®, SYSTAT®, KWIKSTAT®, STATA®, and STATMOST®. (Internet addresses for all these programs are given in appendix F.) You can also take a

random sample with a table of random numbers, like the one in appendix A, taken from the Rand Corporation's volume called *A Million Random Digits with 100,000 Normal Deviates* (1965). The book has no plot or characters, just a million random numbers—a few of which have been reprinted in appendix A.

Just enter the table anywhere. Since the numbers are random, it makes no difference where you start, so long as you don't always start at the same place. Read down a column or across a row. The numbers are in groups of five, in case you ever want to take samples up to 99,999 units. If you are sampling from fewer than 10 units in a population, then look just at the first digit in each group. If you are sampling from a population of 10 to 99 units, then look just at the first two digits, and so on.

Throw out any numbers that are too big for your sample. Say you are taking 300 sample minutes from a population of 5,040 daylight minutes in a week during November in Atlanta, Georgia. (You might do this if you were trying to describe what a family did during that week.) Any number larger than 5,040 is automatically ignored. Just go on to the next number in the table. Ignore duplicate numbers, too.

If you go through the table once (down all the columns) and still don't have enough numbers for your sample then go through it again, starting with the second digit in each group, and then the third. If you began by taking numbers in the columns, take them from rows. You probably won't run out of random numbers for rough-and-ready samples if you use appendix A for the rest of your life.

When you have your list of random numbers, then whoever goes with each one is in the sample. Period. If there are 1,230 people in the population, and your list of random numbers says that you have to interview person #212, then do it. No fair leaving out some people because they are members of the elite and probably wouldn't want to give you the time of day. No fair leaving out people you don't like or don't want to work with. None of that.

In the real world of research, of course, random samples are tampered with all the time. (And no snickering here about the "real world" of research. Social research—in universities, in marketing firms, in polling firms, in the military—is a multibillion-dollar-a-year industry in the United States alone—and that's real enough for most people.) A common form of meddling with samples is when door-to-door interviewers find a sample selectee not at home and go to the nearest house for a replacement. This can have dramatically bad results.

Suppose you go out to interview between 10 A.M. and 4 P.M. People who are home during these hours tend to be old, or sick, or mothers with small children. Of course, those same people are home in the evening, too, but now

they're joined by all the single people home from work, so the average family size goes up.

Telephone survey researchers call back from three to 10 times before replacing a member of a sample. When survey researchers suspect (from prior work) that, say, 25% of a sample won't be reachable, even after call-backs, they increase their original sample size by 25% so the final sample will be both the right size and representative.

Systematic Random Sampling

Most people don't actually do **simple random sampling** these days; instead they do something called **systematic random sampling** because it is much, much easier to do. If you are dealing with an unnumbered sampling frame of 48,673 (the student population at the University of Florida in 2003), then simple random sampling is nearly impossible. You would have to number all those names first. In doing systematic random sampling, you need a random start and a **sampling interval**, N . You enter the sampling frame at a randomly selected spot (using appendix A again) and take every N th person (or item) in the frame.

In choosing a random start, you only need to find one random number in your sampling frame. This is usually easy to do. If you are dealing with 48,673 names, listed on a computer printout, at 400 to a page, then number 9,457 is 257 names down from the top of page 24.

The sampling interval depends on the size of the population and the number of units in your sample. If there are 10,000 people in the population, and you are sampling 400 of them, then after you enter the sampling frame (the list of 10,000 names) you need to take every 25th person ($400 \times 25 = 10,000$) to ensure that every person has at least one chance of being chosen. If there are 640 people in a population, and you are sampling 200 of them, then you would take every 4th person. If you get to the end of the list and you are at number 2 in an interval of 4, just go to the top of the list, start at 3, and keep on going.

Periodicity and Systematic Sampling

I said that systematic sampling *usually* produces a representative sample. When you do systematic random sampling, be aware of the **periodicity** problem. Suppose you're studying a big retirement community in South Florida. The development has 30 identical buildings. Each has six floors, with 10 apartments on each floor, for a total of 1,800 apartments. Now suppose that each floor has one big corner apartment that costs more than the others and

attracts a slightly more affluent group of buyers. If you do a systematic sample of every 10th apartment then, depending on where you entered the list of apartments, you'd have a sample of 180 corner apartments or no corner apartments at all.

David and Mary Hatch (1947) studied the Sunday society pages of the *New York Times* for the years 1932–1942. They found only stories about weddings of Protestants and concluded that the elite of New York must therefore be Protestant. Cahnman (1948) pointed out that the Hatches had studied only June issues of the *Times*. It seemed reasonable. After all, aren't most society weddings in June? Well, yes. Protestant weddings. Upper-class Jews married in other months, and the *Times* covered those weddings as well.

You can avoid the periodicity problem by doing simple random sampling, but if that's not possible, another solution is to make two systematic passes through the population using different sampling intervals. Then you can compare the two samples. Any differences should be attributable to sampling error. If they're not, then you might have a periodicity problem.

Sampling from a Telephone Book

Systematic sampling is fine if you know that the sampling frame has 48,673 elements. What do you do when the size of the sampling frame is unknown? A big telephone book is an unnumbered sampling frame of unknown size. To use this kind of sampling frame, first determine the number of pages that actually contain listings. To do this, jot down the number of the first and last pages on which listings appear. Most phone books begin with a lot of pages that do not contain listings.

Suppose the listings begin on page 30 and end on page 520. Subtract 30 from 520 and add 1 ($520 - 30 + 1 = 491$) to calculate the number of pages that carry listings.

Then note the number of columns per page and the number of lines per column (count all the lines in a column, even the blank ones).

Suppose the phone book has three columns and 96 lines per column (this is quite typical). To take a random sample of 200 nonbusiness listings from this phone book, take a random sample of 400 page numbers (yes, 400) out of the 491 page numbers between 30 and 520. Just think of the pages as a numbered sampling frame of 491 elements. Next, take a sample of 400 column numbers. Since there are three columns, you want 400 random choices of the numbers 1, 2, 3. Finally, take a sample of 400 line numbers. Since there are 96 lines, you want 400 random numbers between 1 and 96.

Match up the three sets of numbers and pick the sample of listings in the phone book. If the first random number between 30 and 520 is 116, go to page

116. If the first random number between 1 and 3 is 3, go to column 3. If the first random number between 1 and 96 is 43, count down 43 lines. Decide if the listing is eligible. It may be a blank line or a business. That’s why you generate 400 sets of numbers to get 200 good listings.

Telephone books don’t actually make good sampling frames—too many people have unlisted numbers (which is why we have random digit dialing—see chapter 10). But since everyone knows what a phone book looks like, it makes a good example for learning how to sample big, unnumbered lists of things, like the list of Catholic priests in Paraguay or the list of orthopedic surgeons in California.

Stratified Sampling

Stratified random sampling ensures that key subpopulations are included in your sample. You divide a population (a sampling frame) into subpopulations (subframes), based on key independent variables and then take a random (unbiased), sample from each of those subpopulations. You might divide the population into men and women, or into rural and urban subframes—or into key age groups (18–34, 35–49, etc.) or key income groups. As the main sampling frame gets divided by key *independent* variables, the subframes presumably get more and more homogeneous with regard to the key *dependent* variable in the study.

In 1996, for example, representative samples of adult voters in the United States were asked the following question:

Which comes closest to your position? Abortion should be . . .

Legal in all cases	Legal in most cases	Illegal in most cases	Illegal in all cases
-----------------------	------------------------	--------------------------	-------------------------

Across all voters, 60% said that abortion should be legal in all (25%) or most (35%) cases and only 36% said it should be illegal in all (12%) or most (24%) cases. (The remaining 4% had no opinion.)

These facts hide some important differences across religious, ethnic, gender, political, and age groups. Among Catholic voters, 59% said that abortion should be legal in all (22%) or most (37%) cases; among Jewish voters, 91% said that abortion should be legal in all (51%) or most (40%) cases. Among registered Democrats, 72% favored legal abortion in all or most cases; among registered Republicans, 45% took that position (Ladd and Bowman 1997:44–46). Sampling from smaller chunks (by age, gender, and so on) ensures not

only that you capture the variation, but that you also wind up understanding how that variation is distributed.

This is called **maximizing the between-group variance** and **minimizing the within-group variance** for the independent variables in a study. *It's what you want to do in building a sample* because it reduces sampling error and thus makes samples more precise.

This sounds like a great thing to do, but you have to *know what the key independent variables are*. Shoe size is almost certainly not related to what people think is the ideal number of children to have. Gender and generation, however, seem like plausible variables on which to stratify a sample. So, if you are taking a poll to find out the ideal number of children, you might divide the adult population into, say, four generations: 15–29, 30–44, 45–59, and over 59.

With two genders, this creates a **sampling design** with eight strata: men 15–29, 30–44, 45–59, and over 59; women 15–29, 30–44, 45–59, and over 59. Then you take a random sample of people from each of the eight strata and run your poll. If your hunch about the importance of gender and generation is correct, you'll find the attitudes of men and the attitudes of women more homogeneous than the attitudes of men and women thrown together.

Table 6.1 shows the distribution of gender and **age cohorts** for St. Lucia in 2001. The numbers in parentheses are percentages of the total population 15 and older (106,479), not percentages of the column totals.

TABLE 6.1
Estimated Population by Sex and Age Groups for St. Lucia, 2001

<i>Age cohort</i>	<i>Males</i>	<i>Females</i>	<i>Total</i>
15–29	21,097 (19.8%)	22,177 (20.8%)	43,274 (40.6%)
30–44	15,858 (14.9%)	16,763 (15.7%)	32,621 (30.6%)
45–59	8,269 (7.8%)	8,351 (7.8%)	16,620 (15.6%)
>59	7,407 (7%)	6,557 (6.2%)	13,964 (13.1%)
Total	52,631 (49.5%)	53,848 (50.5%)	106,479 (100%)

SOURCE: Govt. Statistics, St. Lucia. <http://www.stats.gov.lc/pop23.htm>.

A **proportionate stratified random sample** of 800 respondents would include 112 men between the ages of 30 and 44 (14% of 800 = 112), but 120 women between the ages of 30 and 44 (15% of 800 = 120), and so on.

Watch out, though. We're asking people about their ideal family size and thinking about stratifying by gender because we're accustomed to thinking in terms of gender on questions about family size. But gender-associated preferences are changing rapidly in late industrial societies, and we might be way

off base in our thinking. Separating the population into gender strata might just be creating unnecessary work. Worse, it might introduce unknown error. If your guess about age and gender being related to desired number of children is wrong, then using table 6.1 to create a sampling design will just make it harder for you to discover your error.

Here are the rules on stratification: (1) If differences on a dependent variable are large across strata like age, sex, ethnic group, and so on, then stratifying a sample is a great idea. (2) If differences are small, then stratifying just adds unnecessary work. (3) If you are uncertain about the independent variables that could be at work in affecting your dependent variable, then leave well enough alone and don't stratify the sample. *You can always stratify the data* you collect and test various stratification schemes in the analysis instead of in the sampling.

Disproportionate Sampling

Disproportionate stratified random sampling is appropriate whenever an important subpopulation is likely to be underrepresented in a simple random sample or in a stratified random sample. Suppose you are doing a study of factors affecting grade-point averages among college students. You suspect that the independent variable called "race" has some effect on the dependent variable.

Suppose further that 5% of the student population is African American and that you have time and money to interview 400 students out of a population of 8,000. If you took 10,000 samples of 400 each from the population (replacing the 400 each time, of course), then the average number of African Americans in all the samples would approach 20—that is, 5% of the sample.

But you are going to take *one* sample of 400. It might contain exactly 20 (5%) African Americans; on the other hand, it might contain just 5 (1.25%) African Americans. To ensure that you have enough data on African American students and on white students, you put the African Americans and the Whites into separate *strata* and draw two random samples of 200 each. The African Americans are disproportionately sampled by a factor of 10 (200 instead of the expected 20).

Native Americans comprise just 8/10 of 1% of the population of the United States. If you take a thousand samples of 1,000 Americans at random, you expect to run into about eight Native Americans, on average, across all the samples. (Some samples will have no Native Americans, and some may have 20, but, on average, you'll get about eight.) Without disproportionate sampling, Native Americans would be underrepresented in any national survey in the United States. When Sugarman et al. (1994) ran the National Maternal and

Infant Health Survey in 1988, they used birth certificates dated July 1–December 31, 1988, from 35 Native American health centers as their sampling frame and selected 1,480 eligible mothers for the study of maternal and infant health in that population.

Weighting Results

One popular method for collecting data about daily activities is called **experience sampling** (Csikszentmihalyi and Larson 1987). You give a sample of people a beeper. They carry it around and you beep them at random times during the day. They fill out a little form about what they're doing at the time. (We'll look more closely at this method in chapter 15).

Suppose you want to contrast what people do on weekends and what they do during the week. If you beep people, say, eight times during each day, you'll wind up with 40 reports for each person for the 5-day workweek but only 16 forms for each person for each 2-day weekend because you've sampled the two strata—weekdays and weekends—proportionately.

If you want more data points for the weekend, you might beep people 12 times on Saturday and 12 times on Sunday. That gives you 24 data points, but you've disproportionately sampled one stratum. The weekend represents $2/7$, or 28.6% of the week, but you've got 64 data points and 24 of them, or 37.5%, are about the weekend. Before comparing any data across the strata, you need to make the weekend data and the weekday data statistically comparable.

This is where **weighting** comes in. Multiply each weekday data point by 1.50 so that the 40 data points become worth 60 and the 24 weekend data points are again worth exactly $2/7$ of the total.

You should also weight your data when you have unequal response rates in a stratified sample. Suppose you sample 200 farmers and 200 townspeople in a rural African district where 60% of the families are farmers and 40% are residents of the local town. Of the 400 potential informants, 178 farmers and 163 townspeople respond to your questions. If you compare the answers of farmers and townspeople on a variable, you'll need to weight each farmer's data by $178/163 = 1.09$ times each townsperson's data on that variable. That takes care of the unequal response rates.

Then, you'll need to weight each farmer's data as counting 1.5 times each townsperson's data on the variable. That takes care of the fact that there are half again as many farmers as there are people living in town. This seems complicated because it is. In the BC era (before computers), researchers had to work very hard to use disproportionate sampling. Fortunately, these days weighting is a simple procedure available in all major statistical analysis packages.

Cluster Sampling and Complex Sampling Designs

Cluster sampling is a way to sample populations for which there are no convenient lists or frames. It's also a way to minimize travel time in reaching scattered units of data collection. Cluster sampling is based on the fact that people act out their lives in more or less natural groups, or "clusters." They live in geographic areas (like counties, precincts, states, and so on), and they participate in the activities of institutions (like schools, churches, brotherhoods, credit unions, and so on). Even if there are no lists of people whom you want to study, you can sample areas or institutions and locate a sample within those clusters.

For example, there are no lists of schoolchildren in large cities, but children cluster in schools. There *are* lists of schools, so you can take a sample of them, and then sample children within each school selected. The idea in cluster sampling is to narrow the sampling field down from large, heterogeneous chunks to small, homogeneous ones that are relatively easy to sample directly.

Sometimes, though, you have to create the initial sampling frame on your own. In chapter 5, I mentioned a study in which Lambros Comitas and I compared Greeks who had returned from West Germany as labor migrants with Greeks who had never left their country (Bernard and Comitas 1978). There were no lists of returned migrants, so we decided to locate the children of returned migrants in the Athens schools and use them to select a sample of their parents. The problem was, we couldn't get a list of schools in Athens.

So we took a map of the city and divided it into small bits by laying a grid over it. Then we took a random sample of the bits and sent interviewers to find the school nearest each bit selected. The interviewers asked the principal of each school to identify the children of returned labor migrants. (It was easy for the principal to do, by the way. The principal said that all the returned migrant children spoke Greek with a German accent.) That way, we were able to make up two lists for each school: one of children who had been abroad, and one of children who had not. By sampling children randomly from those lists at each school, we were able to select a representative sample of parents.

This two-stage sampling design combined a cluster sample with a simple random sample to select the eventual units of analysis.

Laurent et al. (2003) wanted to assess the rate of sexually transmitted diseases among unregistered female sex workers in Dakar, Senegal. Now, by definition, unregistered means no list, so the researchers used a two-stage cluster sample. They created a sampling frame of all registered and all clandestine bars in Dakar, plus all the unregistered brothels, and all the nightclubs. They did this over a period of several months with the help of some women prostitutes, some local physicians who had treated female sex workers, the police,

and two social workers, each of whom had worked with female sex workers for over 25 years. Laurent et al. calculated that they needed 94 establishments, so they chose a simple random sample of places from the list of 183. Then they went in teams to each of the 94 places and interviewed all the unregistered prostitutes who were there at the time of the visit. The study by Laurent et al. combined a random sample of clusters with an opportunity sample of people.

Anthony and Suely Anderson (1983) wanted to compare people in Bacabal County, Brazil, who exploited the babassu palm with those who didn't. There was no list of households, but they did manage to get a list of the 344 named hamlets in the county. They divided the hamlets into those that supplied whole babassu fruits to new industries in the area and those that did not. Only 10.5% of the 344 hamlets supplied fruits to the industries, so the Andersons selected 10 hamlets randomly from each group for their survey. In other words, in the first stage of the process they stratified the clusters and took a disproportionate random sample from one of the clusters.

Next, they did a census of the 20 hamlets, collecting information on every household and particularly whether the household had land or was landless. At this stage, then, they created a sampling frame (the census) and stratified the frame into landowning and landless households. Finally, they selected 89 landless households randomly for interviewing. This was 25% of the stratum of landless peasants. Since there were only 61 landowners, they decided to interview the entire population of this stratum.

Sampling designs can involve several stages. If you are studying Haitian refugee children in Miami, you could take a random sample of schools, but if you do that, you'll almost certainly select some schools in which there are no Haitian children. A three-stage sampling design is called for.

In the first stage, you would make a list of the neighborhoods in the city, find out which ones are home to a lot of refugees from Haiti, and sample those districts. In the second stage, you would take a random sample of schools from each of the chosen districts. Finally, in the third stage, you would develop a list of Haitian refugee children in each school and draw your final sample.

Al-Nuaim et al. (1997) used multistage stratified cluster sampling in their national study of adult obesity in Saudi Arabia. In the first stage, they selected cities and villages from each region of the country so that each region's total population was proportionately represented. Then they randomly selected districts from the local maps of the cities and villages in their sample. Next, they listed all the streets in each of the districts and selected every third street. Then they chose every third house on each of the streets and asked each adult in the selected houses to participate in the study.

Probability Proportionate to Size

The best estimates of a parameter are produced in samples taken from clusters of equal size. When clusters are not equal in size, then samples should be taken **PPS**—with **probability proportionate to size**.

Suppose you had money and time to do 800 household interviews in a city of 50,000 households. You intend to select 40 blocks, out of a total of 280, and do 20 interviews in each block. You want each of the 800 households in the final sample to have exactly the same probability of being selected.

Should each block be equally likely to be chosen for your sample? No, because census blocks never contribute equally to the total population from which you will take your final sample. A block that has 100 households in it *should* have twice the chance of being chosen for 20 interviews as a block that has 50 households and half the chance of a block that has 200 households.

When you get down to the block level, each household on a block with 100 residences has a 20% (20/100) chance of being selected for the sample; each household on a block with 300 residences has only a 6.7% (20/300) chance of being selected.

Lené Levy-Storms (1998, n.d.) wanted to talk to older Samoan women in Los Angeles County about mammography. The problem was not that women were reticent to talk about the subject. The problem was: How do you find a representative sample of older Samoan women in Los Angeles County?

From prior ethnographic research, Levy-Storms knew that Samoan women regularly attend churches where the minister is Samoan. She went to the president of the Samoan Federation of America in Carson, California, and he suggested nine cities in L.A. County where Samoans were concentrated. There were 60 churches with Samoan ministers in the nine cities, representing nine denominations. Levy-Storms asked each of the ministers to estimate the number of female church members who were over 50. Based on these estimates, she chose a PPS sample of 40 churches (so that churches with more or fewer older women were properly represented). This gave her a sample of 299 Samoan women over 50. This clever sampling strategy really worked: Levy-Storms contacted the 299 women and wound up with 290 interviews—a 97% cooperation rate.

PPS sampling is called for under three conditions: (1) when you are dealing with large, unevenly distributed populations (such as cities that have high-rise and single-family neighborhoods); (2) when your sample is large enough to withstand being broken up into a lot of pieces (clusters) without substantially increasing the sampling error; and (3) when you have data on the population

of many small blocks in a population and can calculate their respective proportionate contributions to the total population.

PPS Samples in the Field

What do you do when you don't have neat clusters and neat sampling frames printed out on a computer by a reliable government agency? The answer is to place your trust in randomness and *create* maximally heterogeneous clusters from which to take a random sample.

Draw or get a map of the area you are studying. Place 100 numbered dots around the edge of the map. Try to space the numbers equidistant from one another, but don't worry if they are not. Select a pair of numbers at random and draw a line between them. Now select another pair of numbers (replace the first pair before selecting the second) and draw a line between them. In the unlikely event that you choose the same pair twice, simply choose a third pair. Keep doing this, replacing the numbers each time. After you've drawn 20–50 lines across the map (depending on the size of the map), you can begin sampling.

Notice that the lines drawn across the map (see figure 6.1) create a lot of wildly uneven spaces. Since you don't know the distribution of population density in the area you are studying, this technique maximizes the chance that

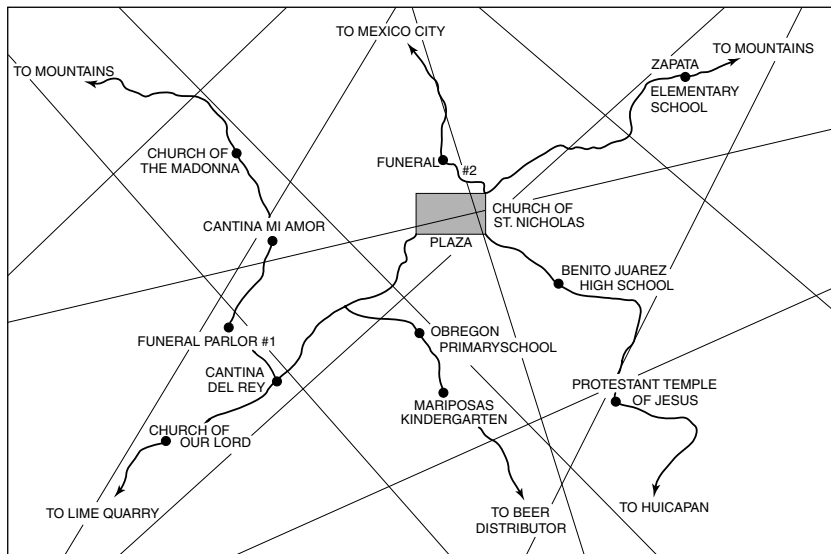


Figure 6.1. Creating maximally heterogeneous sampling clusters in the field.

you will properly survey the population, more or less PPS. By creating a series of (essentially) random chunks of different sizes, you distribute the error you might introduce by not knowing the density, and that distribution lowers the possible error.

Number the uneven spaces created by the lines and choose some of them at random. Go to those spaces, number the households, and select an appropriate number at random. Remember, you want to have the same number of households from *each* made-up geographic cluster, no matter what its size. If you are doing 400 interviews, you would select 20 geographic chunks and do 20 interviews or behavioral observations in each.

My colleagues and I used this method in 1986 to find out how many people in Mexico City knew someone who died in that city's monster earthquake the year before (Bernard et al. 1989). Instead of selecting households, though, our interview team went to each geographic chunk we'd selected and stopped the first 10 people they ran into on the street at each point. This is called a **street-intercept survey**.

Miller et al. (1997) sampled blocks of streets in a city and did a street-intercept survey of African American men. They compared the results to a random-digit dialing telephone survey in the same city. The street-intercept survey did a better job of representing the population than did the telephone survey. For one thing, the response rate for the street intercept survey was over 80%. Compare that to the typical telephone survey, where half or more respondents may refuse to be interviewed. Also, with telephone surveys, the socioeconomic profile of respondents is generally higher than in the population (partly because more affluent people agree more often to be interviewed).

A variant of the street-intercept method, called the **mall-intercept** survey, is used widely in market research. Note, though, that the mall-intercept technique is based on quota sampling, not random, representative sampling. (For examples of mall intercept studies, see Gates and Solomon 1982, Bush and Hair 1985, and Hornik and Ellis 1988.)

Penn Handwerker (1993) used a map-sampling method in his study of sexual behavior on Barbados. In his variation of map sampling, you generate 10 random numbers between 0 and 360 (the degrees on a compass). Next, put a dot in the center of a map that you will use for the sampling exercise, and use a protractor to identify the 10 randomly chosen compass points. You then draw lines from the dot in the center of the map through all 10 points to the edge of the map and interview people (or observe houses, or whatever) along those lines. (See Duranleau [1999] for an empirical test of the power of map sampling.)

If you use this technique, you may want to establish a sampling interval (like every fifth case, beginning with the third case). If you finish interviewing

along the lines and don't have enough cases, you can take another random start, with the same or a different interval and start again. Be careful of periodicity, though.

Camilla Harshbarger (1995) used another variation of map sampling in her study of farmers in North West Province, Cameroon (1993). To create a sample of 400 farmers, she took a map of a rural community and drew 100 dots around the perimeter. She used a random number table to select 50 pairs of dots and drew lines between them. She numbered the points created by the crossing of lines, and chose 80 of those points at random. Then, Harshbarger and her field assistants interviewed one farmer in each of the five compounds they found closest to each of the 80 selected dots. (If you use this dot technique, remember to include the points along the edges of the map in your sample, or you'll miss households on those edges.)

There are times when a random, representative sample is out of the question. After Harshbarger did those interviews with 400 randomly selected farmers in North West Province, Cameroon, she set out to interview Fulani cattle herders in the same area. Here's what Harshbarger wrote about her experience in trying to interview the herders:

It was rainy season in Wum and the roads were a nightmare. The grazers lived very far out of town and quite honestly, my research assistants were not willing to trek to the compounds because it would have taken too much time and we would never have finished the job. I consulted X and he agreed to call selected people to designated school houses on certain days. We each took a room and administered the survey with each individual grazer.

Not everyone who was called came for the interview, so we ended up taking who we could get. Therefore, the Wum grazer sample was not representative and initially that was extremely difficult for me to accept. Our team had just finished the 400-farmer survey of Wum that *was* representative, and after all that work it hit me hard that the grazer survey would not be. To get a representative sample, I would have needed a four-wheel drive vehicle, a driver, and more money to pay research assistants for a lengthy stay in the field. Eventually, I forgave myself for the imperfection. (personal communication)

The lessons here are clear. (1) If you are ever in Harshbarger's situation, you, too, can forgive yourself for having a nonrepresentative sample. (2) Even then, like Harshbarger, you should feel badly about it. (For more on space sampling, see Daley et al. 2001 and Lang et al. 2004.)

Maximizing Between-Group Variance: The Wichita Study

Whenever you do multistage cluster sampling, be sure to take as large a sample as possible from the largest, most heterogeneous clusters. The larger

the cluster, the larger the **between-group variance**; the smaller the cluster, the smaller the between-group variance. Counties in the United States are more like each other on any variable (income, race, average age, whatever) than states are; towns within a county are more like each other than counties are; neighborhoods in a town are more like each other than towns are; blocks are more like each other than neighborhoods are. In sampling, the rule is: *maximize between-group variance*.

What does this mean in practice? Following is an actual example of multi-stage sampling from John Hartman's study of Wichita, Kansas (Hartman 1978; Hartman and Hedblom 1979:160ff.). At the time of the study, in the mid-1970s, Wichita had a population of about 193,000 persons over 16. This was the population to which the study team wanted to generalize. The team decided that they could afford only 500 interviews. There are 82 census tracts in Wichita from which they randomly selected 20. These 20 tracts then became the actual population of their study. We'll see in a moment how well their actual study population simulated (represented) the study population to which they wanted to generalize.

Hartman and Hedblom added up the total population in the 20 tracts and divided the population of *each tract* by the total. This gave the percentage of people that each tract, or cluster, contributed to the new population total. Since the researchers were going to do 500 interviews, each tract was assigned that percentage of the interviews. If there were 50,000 people in the 20 tracts, and one of the tracts had a population of 5,000, or 10% of the total, then 50 interviews (10% of the 500) would be done in that tract.

Next, the team numbered the blocks in each tract and selected blocks at random until they had enough for the number of interviews that were to be conducted in that tract. When a block was selected it stayed in the pool, so in some cases more than one interview was to be conducted in a single block. This did not happen very often, and the team wisely left it up to chance to determine this.

This study team made some excellent decisions that maximized the heterogeneity (and hence the representativeness) of their sample. As clusters get smaller and smaller (as you go from tract to block to household, or from village to neighborhood to household), the homogeneity of the units of analysis within the clusters gets greater and greater. People in one census tract or village are more like each other than people in different tracts or villages. People in one census block or barrio are more like each other than people across blocks or barrios. And people in households are more like each other than people in households across the street or over the hill.

This is very important. Most researchers would have no difficulty with the idea that they should only interview one person in a household because, for

example, husbands and wives often have similar ideas about things and report similar behavior with regard to kinship, visiting, health care, child care, and consumption of goods and services. Somehow, the lesson becomes less clear when new researchers move into clusters that are larger than households.

But the rule stands: Maximize heterogeneity of the sample by taking as many of the biggest clusters in your sample as you can, and as many of the next biggest, and so on, always at the expense of the number of clusters at the bottom where homogeneity is greatest. Take more tracts or villages, and fewer blocks per tract or barrios per village. Take more blocks per tract or barrios per village, and fewer households per block or barrio. Take more households and fewer persons per household.

Many survey researchers say that, as a rule, you should have no fewer than five households in a census block. The Wichita group did not follow this rule; they only had enough money and person power to do 500 interviews and they wanted to maximize the likelihood that their sample would represent faithfully the characteristics of the 193,000 adults in their city.

The Wichita team drew two samples—one main sample and one alternate sample. Whenever they could not get someone on the main sample, they took the alternate. That way, they maximized the representativeness of their sample because the alternates were chosen with the same randomized procedure as the main respondents in their survey. They were not forced to take “next door-neighbors” when a main respondent wasn’t home. (This kind of winging it in survey research has a tendency to clobber the representativeness of samples. In the United States, at least, interviewing only people who are at home during the day produces results that represent women with small children, shut-ins, telecommuters, and the elderly—and not much else.)

Next, the Wichita team randomly selected the households for interview within each block. This was the third stage in this multistage cluster design. The fourth stage consisted of flipping a coin to decide whether to interview a man or a woman in households with both. Whoever came to the door was asked to provide a list of those in the household over 16 years of age. If there was more than one eligible person in the household, the interviewer selected one at random, conforming to the decision made earlier on sex of respondent.

Table 6.2 shows how well the Wichita team did.

All in all, they did very well. In addition to the variables shown in the table here, the Wichita sample was a fair representation of marital status, occupation, and education, although on this last independent variable there were some pretty large discrepancies. For example, according to the 1970 census, 8% of the population of Wichita had less than 8 years of schooling, but only 4% of the sample had this characteristic. Only 14% of the general population

TABLE 6.2
Comparison of Survey Results and Population Parameters for the
Wichita Study by Hartman and Hedblom

	<i>Wichita in 1973</i>	<i>Hartman and Hedblom's Sample for 1973</i>
White	86.8%	82.8%
African	9.7%	10.8%
Chicano	2.5%	2.6%
Other	1.0%	2.8%
Male	46.6%	46.9%
Female	53.4%	53.1%
Median age	38.5	39.5

SOURCE: *Methods for the Social Sciences: A Handbook for Students and Non-Specialists*, by J. J. Hartman and J. H. Hedblom, 1979, p. 165. Copyright © 1979 John J. Hartman and Jack H. Hedblom. Reproduced with permission of Greenwood Publishing Group, Inc., Westport, CT.

had completed from 1 to 2 years of college, but 22% of the sample had that much education.

All things considered, though, the sampling procedure followed in the Wichita study was a model of technique, and the results show it. Whatever they found out about the 500 people they interviewed, the researchers could be very confident that the results were generalizable to the 193,000 adults in Wichita.

In sum: If you don't have a sampling frame for a population, try to do a multistage cluster sample, narrowing down to natural clusters that do have lists. Sample heavier at the higher levels in a multistage sample and lighter at the lower stages.

Just in case you're wondering if you can do this under difficult field conditions, Oyuela-Caycedo and Vieco Albarracín (1999) studied the social organization of the Ticuna Indians of the Colombian Amazon. Most of the 9,500 Ticuna in Colombia are in 32 hamlets, along the Amazon, the Loreta Yacu, the Cotuhé, and the Putumayo Rivers. The Ticuna live in large houses that comprise from one to three families, including grandparents, unmarried children, and married sons with their wives and children. To get a representative sample of the Ticuna, Oyuela-Caycedo and Vieco Albarracín selected six of the 32 hamlets along the four rivers and made a list of the household heads in those hamlets. Then, they numbered the household heads and, using a table of random numbers (like the one in appendix A), they selected 50 women and 58 men. Oyuela-Caycedo and Vieco Albarracín had to visit some of the selected houses several times in order to secure an interview, but they wound up interviewing all the members of their sample.

Is the sample representative? We can't know for sure, but take a look at

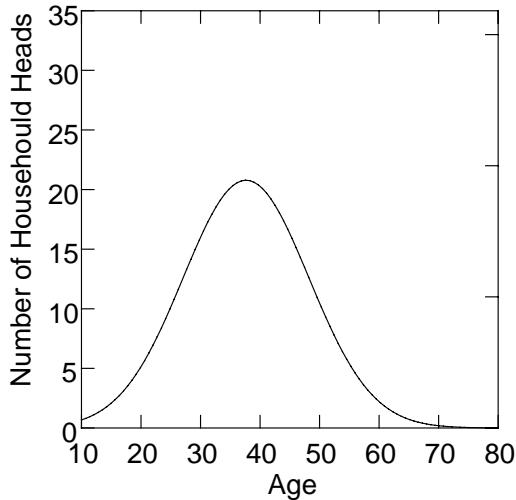


Figure 6.2a. Distribution of the ages of Ticuna household heads.

SOURCE: Adapted from data in A. Oyuela-Caycedo and J. J. Vieco Albarracín, “Aproximación cuantitativa a la organización social de los ticuna del trapecio amazónico colombiano,” *Revista Colombiana de Antropología*, Vol. 35, pp. 146–79, figure 1, p. 157, 1999.

figure 6.2. Figure 6.2a shows the distribution of the ages of the household heads in the sample; figure 6.2b shows the distribution of the number of children in the households of the sample. Both curves look very normal—just what we expect from variables like age and number of children (more about normal distributions in chapter 7). If the sample of Ticuna household heads represents what we expect from age and number of children, then any other variables the research team measured are likely (not guaranteed, just likely) to be representative, too.

How Big Should a Sample Be?

There are two things you can do to get good samples. (1) You can ensure **sample accuracy** by making sure that every element in the population has an equal chance of being selected—that is, you can make sure the sample is unbiased. (2) You can ensure **sample precision** by increasing the size of unbiased samples. We’ve already discussed the importance of how to make samples unbiased. The next step is to decide how big a sample needs to be.

Sample size depends on: (1) the heterogeneity of the population or chunks of population (strata or clusters) from which you choose the elements; (2) how

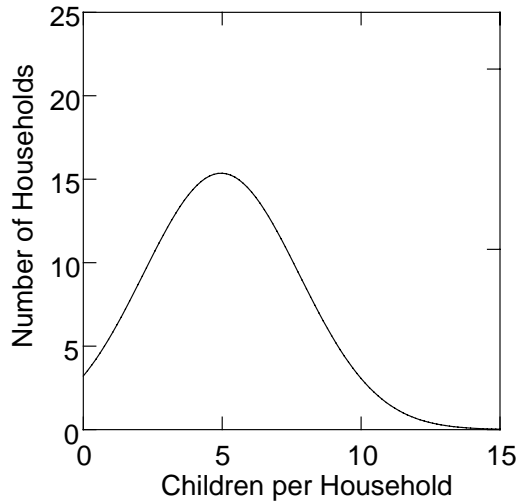


Figure 6.2b. Distribution of the number of children in Ticuna households.

SOURCE: Adapted from data in A. Oyuela-Caycedo and J. J. Vieco Albarracín, “Aproximación cuantitativa a la organización social de los ticuna del trapezio amazónico colombiano,” *Revista Colombiana de Antropología*, Vol. 35, pp. 146–79, table 6, p. 159, 1999.

many population subgroups (that is, independent variables) you want to deal with simultaneously in your analysis; (3) the size of the subgroup that you’re trying to detect; and (4) how precise you want your **sample statistics** (or **parameter estimators**) to be.

1. **Heterogeneity of the population.** When all elements of a population have the same score on some measure, a sample of 1 will do. Ask a lot of people to tell you how many days there are in a week and you’ll soon understand that a big sample isn’t going to uncover a lot of heterogeneity. But if you want to know what the average ideal family size is, you may need to cover a lot of social ground. People of different ethnicities, religions, incomes, genders, and ages may have very different ideas about this. (In fact, these independent variables may interact in complex ways. Multivariate analysis tells you about this interaction. We’ll get to this in chapter 21.)
2. **The number of subgroups in the analysis.** Remember the **factorial design** problem in chapter 5 on experiments? We had three independent variables, each with two attributes, so we needed eight groups ($2^3 = 8$). It wouldn’t do you much good to have, say, one experimental subject in each of those eight groups. If you’re going to analyze all eight of the conditions in the experiment, you’ve got to fill each of the conditions with some reasonable number of subjects. If you

have only 15 people in each of the eight conditions, then you need a sample of 120.

The same principle holds when you're trying to figure out how big a sample you need for a survey. In the example above about stratified random sampling, we had four generations and two genders—which also produced an eight-cell design.

If all you want to know is a single proportion—like what percentage of people in a population approve or disapprove of something—then you need about 100 respondents to be 95% confident, within plus or minus 3 points, that your sample estimate is within 2 standard deviations of the population parameter (more about confidence limits, normal distributions, standard deviations, and parameters coming next). But if you want to know whether women factory workers in Rio de Janeiro who earn less than \$300 per month have different opinions from, say, middle-class women in Rio whose family income is more than \$600 per month, then you'll need a bigger sample.

3. **The size of the subgroup.** If the population you are trying to study is rare and hard to find, and if you have to rely on a simple random sample of the entire population, you'll need a very large initial sample. A needs assessment survey of people over 75 in Florida took 72,000 phone calls to get 1,647 interviews—about 44 calls per interview (Henry 1990:88). This is because only 6.5% of Florida's population was over 75 at the time of the survey. By contrast, the monthly Florida survey of 600 representative consumers takes about 5,000 calls (about eight per interview). That's because just about everyone in the state 18 and older is a consumer and is eligible for the survey. (Christopher McCarty, personal communication)

The smaller the difference on any measure between two populations, the bigger the sample you need in order to detect that difference. Suppose you suspect that Blacks and Whites in a prison system have received different sentences for the same crime. Henry (1990:121) shows that a difference of 16 months in sentence length for the same crime would be detected with a sample of just 30 in each racial group (if the members of the sample were selected randomly, of course). To detect a difference of 3 months, however, you need 775 in each group.

4. **Precision.** This one takes us into sampling theory.

7



Sampling Theory

Distributions

At the end of this chapter, you should understand why it's possible to estimate *very accurately, most of the time*, the average age of the 213 million adults in the United States by talking to just 1,600 of them. And you should understand why you can also do this *pretty accurately, much of the time*, by talking to just 400 people. Sampling theory is partly about **distributions**, which come in a variety of shapes. Figure 7.1 shows what is known as the **normal distribution**

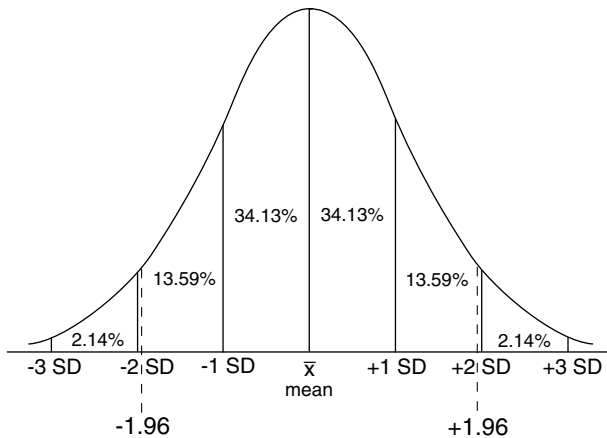


Figure 7.1. The normal curve and the first, second, and third standard deviation.

The Normal Curve and z-Scores

The so-called normal distribution is generated by a formula that can be found in many intro statistics texts. The distribution has a mean of 0 and a **standard deviation** of 1. The standard deviation is a measure of how much the scores in a distribution vary from the mean score. The larger the standard deviation, the more dispersion around the mean. Here's the formula for the standard deviation, or *sd*. (We will take up the *sd* again in chapter 19. The *sd* is the square root of the variance, which we'll take up in chapter 20.)

$$sd = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}} \quad \text{Formula 7.1}$$

The symbol \bar{x} (read: x-bar) in formula 7.1 is used to signify the *mean of a sample*. The mean of a *population* (the parameter we want to estimate), is symbolized by μ (the Greek lower-case letter "mu," pronounced "myoo"). The standard deviation of a population is symbolized by σ (the Greek lower-case letter "sigma"), and the standard deviation of a sample is written as *SD* or *sd* or *s*. The standard deviation is the square root of the sum of all the squared differences between every score in a set of scores and the mean, divided by the number of scores minus 1.

The standard deviation of a sampling distribution of means is the **standard error** of the mean, or SEM. The formula for calculating SEM is:

$$SEM = \frac{sd}{\sqrt{n}} \quad \text{Formula 7.2}$$

where *n* is the sample size. In other words, the standard error of the mean gives us an idea of how much a sample mean varies from the mean of the population that we're trying to estimate.

Suppose that in a sample of 100 merchants in a small town in Malaysia, you find that the average income is RM12,600 (about \$3,300 in 2005 U.S. dollars), with a standard deviation of RM4,000 (RM is the symbol for the Malaysian Ringgit). The standard error of the mean is:

$$12,600 \mp \frac{4,000}{\sqrt{100}} = 12,600 \pm 400$$

Do the calculation:

$$\begin{aligned} 12,600 + 400 &= 13,000 \\ 12,600 - 400 &= 12,200 \end{aligned}$$

In normal distributions—that is, distributions that have a mean of 0 and a standard deviation of 1—exactly 34.135% of the area under the curve (the white space between the curve and the baseline) is contained between the perpendicular line that represents the mean in the middle of the curve and the line that rises from the baseline at 1 standard deviation above and 1 standard deviation below the mean.

Appendix B is a table of **z-scores**, or **standard scores**. These scores are the number of standard deviations from the mean in a normal distribution, in increments of 1/100th of a standard deviation. For each *z*-score, beginning with 0.00 standard deviations (the mean) and on up to 3.09 standard deviations (on either side of the mean), appendix B shows the *percentage of the physical area under the curve of a normal distribution*. That percentage represents the percentage of cases that fall within any number of standard deviations above and below the mean in a normally distributed set of cases.

We see from appendix B that 34.13% of the area under the curve is one standard deviation above the mean and another 34.13% is one standard deviation below the mean. (The reason that this is so is beyond what we can go into here. For more on this, consult a standard statistics text.) Thus, 68.26% of all scores in a normal distribution fall within one standard deviation of the mean. We also see from appendix B that 95.44% of all scores in a normal distribution fall within two standard deviations and that 99.7% fall within three standard deviations.

Look again at figure 7.1. You can see why so many cases are contained within 1 *sd* above and below the mean: The normal curve is tallest and fattest around the mean and much more of the area under the curve is encompassed in the first *sd* from the mean than is encompassed between the first and second *sd* from the mean.

If 95.44% of the area under a normal curve falls within two standard deviations from the mean, then exactly 95% should fall within slightly less than two standard deviations. Appendix B tells us that 1.96 standard deviations above and below the mean account for 95% of all scores in a normal distribution. And, similarly, 2.58 *sd* account for 99% of all scores. This, too, is shown in figure 7.1.

The normal distribution is an idealized form. In practice, many variables are not distributed in the perfectly symmetric shape we see in figure 7.1. Figure 7.2 shows some other shapes for distributions. Figure 7.2a shows a **bimodal distribution**. Suppose the *x*-axis in figure 7.2a is age, and the *y*-axis is strength of agreement, on a scale of 1 to 5, to the question “Did you like the beer commercial shown during the Superbowl yesterday?” The bimodal distribution shows that people in their 20s and people in their 60s liked the commercial, but others didn’t.

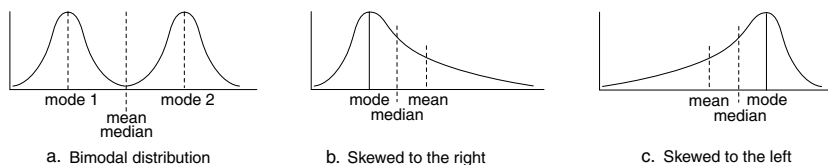


Figure 7.2. Bimodal and skewed distributions.

Figure 7.2b and figure 7.2c are **skewed distributions**. A distribution can be skewed positively (with a long tail going off to the right) or negatively (with the tail going off to the left). Figures 7.2b and 7.2c look like the distributions of scores in two very different university courses. In figure 7.2b, most students got low grades, and there is a long tail of students who got high grades. In figure 7.2c, most students got relatively high grades, and there is a long tail of students who got lower grades.

The normal distribution is symmetric, but not all symmetric distributions are normal. Figure 7.3 shows three variations of a **symmetric distribution**—

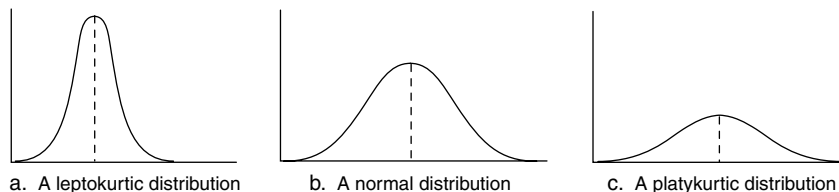


Figure 7.3. Three symmetric distributions including the normal distribution.

that is, distributions for which the mean and the median are the same. The one on the left is **leptokurtic** (from Greek, meaning “thin bulge”) and the one on the right is **platykurtic** (meaning “flat bulge”). The curve in the middle is the famous bell-shaped, normal distribution. In a leptokurtic, symmetric distribution, the standard deviation is less than 1.0; and in a platykurtic, symmetric distribution, the standard deviation is greater than 1.0. The physical distance between marriage partners (whether among tribal people or urbanites) usually forms a leptokurtic distribution. People tend to marry people who live near them, and there are fewer and fewer marriages as the distance between partners increases (Sheets 1982). By contrast, we expect the distribution of height and weight of athletes in the National Basketball Association to be more platykurtic across teams, since coaches are all recruiting players of more-or-less similar build.

The shape of a distribution—normal, skewed, bimodal, and so on—contains

a lot of information about what is going on, but it doesn't tell us *why* things turned out the way they did. A sample with a bimodal or highly skewed distribution is a hint that you might be dealing with more than one population or culture.

The Central Limit Theorem

The fact that many variables are not normally distributed would make sampling a hazardous business, were it not for the **central limit theorem**. According to this theorem, if you take many samples of a population, and *if the samples are big enough*, then:

1. The mean and the standard deviation of the sample means will usually approximate the true mean and standard deviation of the population. (You'll understand why this is so a bit later in the chapter, when we discuss confidence intervals.)
2. The distribution of sample means will approximate a normal distribution.

We can demonstrate both parts of the central limit theorem with some examples.

Part 1 of the Central Limit Theorem

Table 7.1 shows the per capita gross domestic product (PCGDP) for the 50 poorest countries in the world in 1998.

Here is a random sample of five of those countries: Guinea-Bissau, Nepal, Moldova, Zambia, and Haiti. Consider these five as a **population** of units of analysis. In 2000, these countries had an annual per capita GDP, respectively of \$100, \$197, \$374, \$413, and \$443 (U.S. dollars). These five numbers sum to \$1,527 and their average, $1527/5$, is \$305.40.

There are 10 possible samples of two elements in any population of five elements. All 10 samples for the five countries in our example are shown in the left-hand column of table 7.2. The middle column shows the mean for each sample. This list of means is the **sampling distribution**. And the right-hand column shows the **cumulative mean**.

Notice that the mean of the means for all 10 samples of two elements—that is, the mean of the **sampling distribution**—is \$305.40, which is *exactly the actual mean* per capita GDP of the five countries in the population. In fact, it must be: The mean of all possible samples is equal to the parameter that we're trying to estimate.

Figure 7.4a is a **frequency polygon** that shows the distribution of the five actual GDP values. A frequency polygon is just a histogram with lines connecting the tops of the bars so that the shape of the distribution is emphasized.

TABLE 7.1

Per Capita Gross Domestic Product in U.S. Dollars for the 50 Poorest Countries in 1998

<i>Country</i>	<i>PCGDP</i>	<i>Country</i>	<i>PCGDP</i>
Mozambique	92	Comoros	305
Congo	98	Sudan	305
Guinea-Bissau	100	Mauritania	328
Burundi	103	Vietnam	336
Ethiopia	107	Togo	344
Chad	150	Ghana	346
Sierra Leone	154	Uganda	347
Malawi	156	Yemen	354
Niger	159	Gambia	355
Somalia	177	Kyrgyzstan	366
Nepal	197	Kenya	373
Bhutan	199	Moldova	374
Madagascar	208	Equatorial Guinea	377
Eritrea	210	Mongolia	384
Tanzania	213	Benin	399
Tajikistan	219	Zambia	413
Burkina Faso	221	India	422
Rwanda	225	Lesotho	425
Laos	250	North Korea	440
Mali	254	Nicaragua	442
Cambodia	255	Haiti	443
Myanmar	282	Pakistan	458
Liberia	285	Uzbekistan	470
Central African Republic	296	Indonesia	478
Bangladesh	299	Guinea	515

SOURCE: Statistics Division of the United Nations Secretariat and International Labour Office.
<http://www.un.org/depts/unsd/social/inc-eco.htm>

Compare the shape of this distribution to the one in figure 7.4b showing the distribution of the 10 sample means for the five GDP values we're dealing with here. That distribution looks more like the shape of the normal curve: It's got that telltale bulge in the middle.

Part 2 of the Central Limit Theorem

Figure 7.5 shows the distribution of the 50 data points for GDP in table 7.1. The range is quite broad, from \$92 to \$515 per year per person, and the shape of the distribution is more or less normal.

The actual mean of the data in table 7.1 and in figure 7.5—that is, the

TABLE 7.2
All Samples of 2 from 5 Elements

<i>Sample</i>	<i>Mean</i>	<i>Cumulative Mean</i>
Guinea-Bissau and Nepal	$(100 + 197)/2 = 148.5$	148.5
Guinea-Bissau and Moldova	$(100 + 374)/2 = 237.0$	385.5
Guinea-Bissau and Zambia	$(100 + 413)/2 = 256.5$	642.0
Guinea-Bissau and Haiti	$(100 + 443)/2 = 271.5$	913.5
Nepal and Moldova	$(197 + 374)/2 = 285.5$	1199.0
Nepal and Zambia	$(197 + 413)/2 = 305.0$	1504.0
Nepal and Haiti	$(197 + 443)/2 = 320.0$	1824.0
Moldova and Zambia	$(374 + 413)/2 = 393.5$	2217.5
Moldova and Haiti	$(374 + 443)/2 = 408.5$	2626.0
Zambia and Haiti	$(413 + 443)/2 = 428.0$	3054.0
	$\bar{x} = 3054/10 = 305.4$	

parameter we want to estimate—is \$294.16. There are 2,118,760 samples of size 5 that can be taken from 50 elements. I used a random-number generator to select 15 samples of size 5 from the data in table 7.1 and calculated the mean of each sample. Table 7.3 shows the results.

Even in this small set of 15 samples, the mean is \$284.12—quite close to the actual mean of \$294.16. Figure 7.6a shows the distribution of these samples. It has the look of a normal distribution straining to happen. Figure 7.6b shows 50 samples of five from the 50 countries in table 7.1. The strain toward the normal curve is unmistakable and the mean of those 50 samples is \$293.98

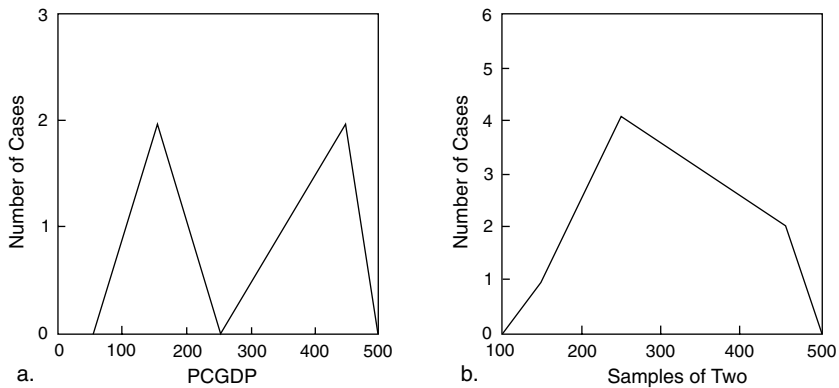


Figure 7.4. Five cases and the distribution of samples of size 2 from those cases.

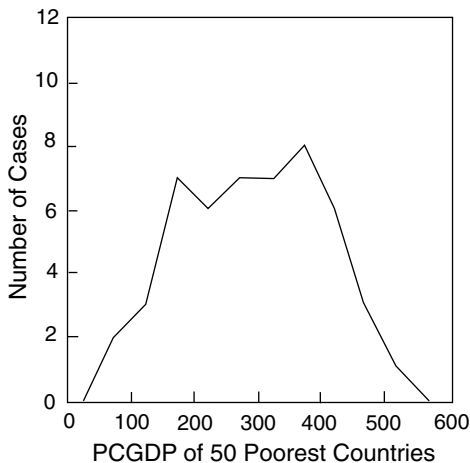


Figure 7.5. The distribution of the 50 data points for GDP in Table 7.1.

(standard deviation 37.44)—very close to the parameter of \$294.16 that we’re trying to estimate.

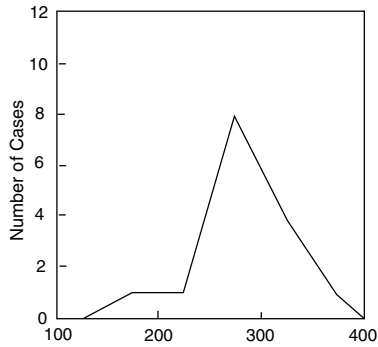
The problem, of course, is that in real research, we don’t get to take 15 or 50 samples. We have to make do with one. The first sample of five elements that I took had a mean of \$305.40—pretty close to the actual mean of \$294.16. But it’s very clear from table 7.3 that any one sample of five elements from table 7.1 could be off by a lot. They range, after all, from \$176.40 to \$342.200. That’s a very big spread, when the real average we’re trying to estimate is \$294.16.

Figure 7.6c shows what happens when we take 30 random samples of size 30 from the 50 elements in table 7.1. Not only is the distribution pretty normal looking, the range of the 30 samples is between \$259.88 and \$321.73, the stan-

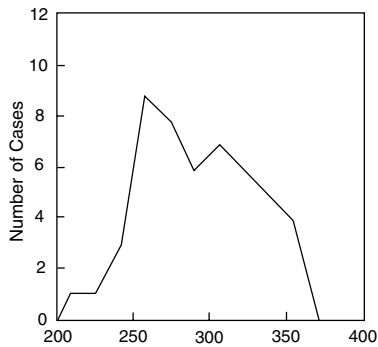
TABLE 7.3
15 Means from Samples of Size 5 Taken from the 50 Elements in Table 7.1

340.00	342.20	319.20
322.00	322.00	254.60
245.00	245.00	329.00
305.40	306.00	205.40
176.40	265.00	283.40

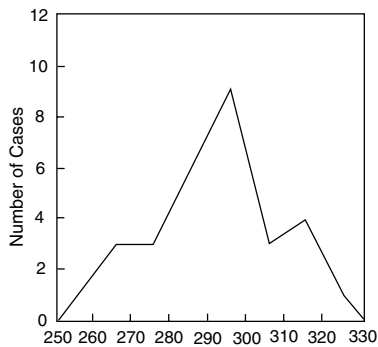
Mean = 284.12
Standard Deviation = 50.49



a. Means for 15 samples of 5 from the 50 GDP values in Table 7.1.



b. Means for 50 samples of 5 from the 50 GDP values in Table 7.1.



c. Means for 30 samples of 30 from the 50 GDP values in Table 7.1.

Figure 7.6. Visualizing the central limit theorem: The distribution of sample means approximates a normal distribution.

standard deviation is down to 16.12, and the mean is \$291.54—very close to the parameter of \$294.16.

We are much closer to answering the question: How big does a sample have to be?

The Standard Error and Confidence Intervals

In the hypothetical example on page 170 we took a sample of 100 merchants in a Malaysian town and found that the mean income was RM12,600, standard error 400. We know from figure 7.1 that 68.26% of all samples of size 100 from this population will produce an estimate that is between 1 standard error above and 1 standard error below the mean—that is, between RM12,200 and RM13,000. The 68.26% **confidence interval**, then, is \$400.

We also know from figure 7.1 that 95.44% of all samples of size 100 will produce an estimate of 2 standard errors, or between RM13,400 and RM11,800. The 95.44% confidence interval, then, is RM800. If we do the sums for the example, we see that the 95% confidence limits are:

$$\text{RM}12,600 \pm 1.96(\text{RM}400) = \text{RM}11,816 \text{ to } \text{RM}13,384$$

and the 99% confidence limits are:

$$\text{RM}12,600 \pm 2.58(\text{RM}400) = \text{RM}11,568 \text{ to } \text{RM}13,632$$

Our “confidence” in these 95% or 99% estimates comes from the power of a random sample and the fact that (by the central limit theorem) sampling distributions are known to be normal irrespective of the distribution of the variable whose mean we are estimating.

What Confidence Limits Are and What They Aren't

If you say that the 95% confidence limits for the estimated mean income are RM11,816 to RM13,384, this does *not* mean that there is a 95% chance that the true mean, μ , lies somewhere in that range. The true mean may or may not lie within that range and we have no way to tell. What we can say, however, is that:

1. If we take a very large number of suitably large random samples from the population (we'll get to what “suitably large” means in a minute); and
2. If we calculate the mean, \bar{x} , and the standard error, SEM, for each sample; and
3. If we then calculate the confidence intervals for each sample mean, based on ± 1.96 SEM; then
4. 95% of these confidence intervals will contain the true mean, μ

Calculating Sample Size for Estimating Means

Now we are *really* close to answering the question about sample size. Suppose we want to get the standard error down to RM200 instead of RM400. We need to solve the following equation:

$$SEM = \frac{sd}{\sqrt{n}} = 4,000/\sqrt{n} = 200$$

Solving for n :

$$\sqrt{n} = \frac{4,000}{200} = 20 \quad \sqrt{n} = 20^2 = 400$$

In other words, to reduce the standard error of the mean from RM400 to RM200, we have to increase the sample size from 100 to 400 people.

Suppose we increase the sample to 400 and we still get a mean of RM12,600 and a standard deviation of RM4,000. The standard error of the mean would then be RM200, and we could estimate, with 95% confidence, that the true mean of the population was between RM12,208 and 12,992. With just 100 people in the sample, the 95% confidence limits were RM11,816 and RM13,384. As the standard error goes down, we get narrower—that is, more precise—confidence limits.

Let's carry this forward another step. If we wanted to get the standard error down to RM100 and the 95% confidence interval down to RM200 from RM400, we would need a sample of 1,600 people. There is a pattern here. To cut the 95% confidence interval *in half*, from RM800 to RM400, we had to *quadruple* the sample size from 100 to 400. To cut the interval *in half again*, to RM200, we'd need to *quadruple* the sample size again, from 400 to 1,600.

There is another pattern, too. If we want to increase our confidence from 95% to 99% that the true mean of the population is within a particular confidence interval, we can raise the multiplier in formula 7.2 from roughly 2 standard deviations to roughly 3. Using the confidence interval of RM400, we would calculate:

$$\sqrt{n} = 3 \left(\frac{4,000}{400} \right) = 30 \quad n = 30^2 = 900$$

We need 900 people, not 400, to be about 99% confident that our sample mean is within RM400, plus or minus, of the parameter.

Small Samples: The t -Distribution

In anthropology, even doing surveys, we often have no choice about the matter and have to use small samples.

What we need is a distribution that is a bit more forgiving than the normal distribution. Fortunately, just such a distribution was discovered by W. S. Gossett, an employee of the Guinness brewery in Ireland. Writing under the pseudonym of “Student,” Gossett described the distribution known as Student’s t . It is based on a distribution that takes into account the fact that statistics calculated on small samples vary more than do statistics calculated on large samples and so have a bigger chance of misestimating the parameter of a continuous variable.

The t distribution is found in appendix C. Figure 7.7 shows graphically the

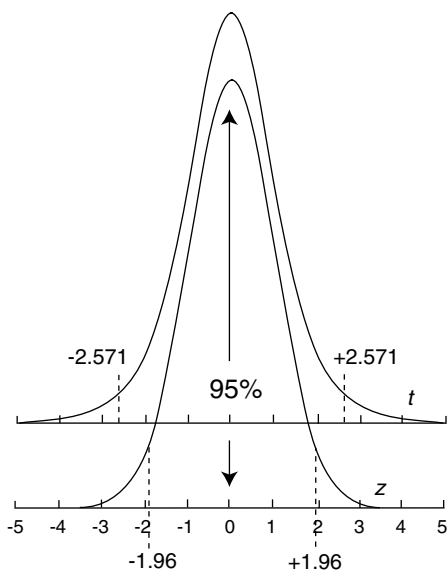


Figure 7.7. Variability in a t distribution and a normal distribution.

difference in the two distributions. In a normal distribution, plus or minus 1.96 sd covers 95% of all sample means. In a t -distribution, with 5 degrees of freedom, it takes $\pm 2.571 sd$ to cover 95% of all sample means.

The confidence interval for small samples, using the t distribution is given in formula 7.3:

$$\text{Confidence Interval} = \bar{x} \pm \left(t_{\alpha/2} \left(\frac{sd}{\sqrt{n}} \right) \right) \quad \text{Formula 7.3}$$

where alpha (α) is the confidence interval you want. If you want to be 95% confident, then $\alpha = .05$. Since half the scores fall above the mean and half below, we divide alpha by two and get .025.

Look up what's called the **critical value** of t in appendix C. In the column for .025, we see that the value is 2.571 with 5 degrees of freedom. Degrees of freedom are one less than the size of the sample, so for a sample of six we need a t statistic of ≥ 2.571 to attain 95% confidence. (The concept of degrees of freedom is described further in chapter 19 in the section on t -tests. And keep in mind that we're interested in the **modulus**, or absolute value of t . A value of -2.571 is just as statistically significant as a value of $+2.571$.)

So, with small samples—which, for practical purposes, means less than 30 units of analysis—we use appendix C (for t) instead of appendix B (for z) to determine the confidence limits around the mean of our estimate. You can see from appendix C that for large samples—30 or more—the difference between the t and the z statistics is negligible. (The t -value of 2.042 for 30 degrees of freedom—which means a sample of 31—is very close to 1.96.)

The Catch

Suppose that instead of estimating the income of a population with a sample of 100, we use a sample of 10 and get the same result—RM12,600 and a standard deviation of RM4,000. For a sample this size, we use the t distribution. With 9 degrees of freedom and an alpha value of .025, we have a t value of 2.262. For a normal curve, 95% of all scores fall within 1.96 standard errors of the mean. The corresponding t value is 2.262 standard errors. Substituting in the formula, we get:

$$\text{RM}12,600 \pm 2.262 \left(\frac{\text{RM}4,000}{\sqrt{10}} \right) = \text{RM}9,739 \text{ to } \text{RM}15,461$$

But there's a catch. With a large sample (greater than 30), we know from the central limit theorem that the sampling distribution will be normal even if the population isn't. Using the t distribution with a small sample, we can calculate the confidence interval around the mean of our sample *only under the assumption that the population is normally distributed*.

In fact, looking back at figure 7.5, we know that the distribution of real data is not perfectly normal. It is somewhat skewed (more about skewed distributions in chapter 19). In real research, we'd never take a sample from a set of just 50 data points—we'd do all our calculations on the full set of the actual data. When we take samples, it's because we don't know what the distribution of the data looks like. And that's why sample size counts.

Calculating Sample Size for Estimating Proportions

And now for proportions. What we've learned so far about estimating the mean of continuous variables (like income and percentages) is applicable to the estimation of proportions as well.

In January 2005, the Gallup Poll reported that 38% of Americans over 18 years of age report keeping one or more guns in their home (down from about 47% in 1959 and from 40% in 1993). The poll included 1,012 respondents and had, as the media say, a “margin of error of plus or minus three percentage points.” This point estimate of 38% means that 385 of the 1,012 people polled said that they had at least one gun in their house.

We can calculate the confidence interval around this point estimate. From the central limit theorem, we know that whatever the true proportion of people is who keep a gun in their home, the estimates of that proportion will be normally distributed if we take a large number of samples of 1,012 people. The formula for determining the 95% confidence limits of a point estimator is:

$$P \text{ (the true proportion)} = \pm 1.96\sqrt{PQ/n} \quad \textbf{Formula 7.4}$$

We use an italicized letter, P , to indicate the true proportion. Our estimate is the regular uppercase P and Q is $1 - P$. Table 7.4 shows what happens to the

TABLE 7.4
Relation of P , Q , and \sqrt{PQ}

<i>If the value of P is really</i>	<i>Then PQ is</i>	<i>and \sqrt{PQ} is</i>
.10 or .90	.09	.30
.20 or .80	.16	.40
.30 or .70	.21	.46
.40 or .60	.24	.49
.50	.25	.50

square root of PQ as the true value of P goes up from 10% to 90% of the population.

We can use our own estimate of P from the Gallup poll in the equation for the confidence limits. Substituting .38 for P and .62 ($1 - .38$) for Q , we get:

$$P = P \pm 1.96 \sqrt{(.38)(.62)/1,012} = .38 \pm .0299$$

which, with rounding, is the familiar “plus or minus three percentage points.” This means that we are 95% confident that the true proportion of adults in the United States who had at least one gun in their home (or at least said they did) at the time this poll was conducted was between 35% and 41%.

Suppose we want to estimate P to within plus or minus 2 percentage points instead of 3 and we still want to maintain the 95% confidence level. We substitute in the formula as follows:

$$P = P \pm 1.96\sqrt{(.38)(.62)/n} = .02$$

and we solve for n :

$$\begin{aligned}n &= 1.96^2(.38)(.62)/.02^2 \\n &= (3.842)(.38)(.62)/.0004 = 2,263\end{aligned}$$

Generalizing, then, the formula for “sample size when estimating proportions in a large population” is:

$$n = z^2(P)(Q)/(\text{confidence interval})^2 \quad \textbf{Formula 7.5}$$

where z is the area under the normal curve that corresponds to the confidence limit we choose. When the confidence limit is 95%, then z is 1.96. When the confidence limit is 99%, then z is 2.58. And so on.

If we start out fresh and have no prior estimate of P , we follow table 7.4 and set P and Q to .5 each. This maximizes the size of the sample for any given confidence interval or confidence level. If we want a sample that produces an estimate of a proportion with a confidence interval of 2 percentage points and we want to be 95% confident in that estimate, we calculate:

$$n \text{ (sample size)} = (1.96)^2(.5)(.5)/(.02)^2 = 2,401$$

In time allocation studies, we estimate the proportion of various behaviors by observing a sample of behaviors. We’ll deal with this in chapter 15, on methods of direct observation (see especially table 15.2).

Estimating Proportions in Samples for Smaller Populations

This general formula, 7.5, is *independent of the size of the population*. Florida has a population of about 17 million. A sample of 400 is .000024 of 17 million; a sample of 2,402 is .00014 of 17 million. Both proportions are microscopic. A sample of 400 from a population of 1 million gets you the same confidence level and the same confidence interval as you get with a sample of 400 from a population of 17 million.

Often, though, we want to take samples from relatively small populations. The key word here is “relatively.” When formula 7.4 or 7.5 calls for a sample that turns out to be 5% or more of the total population, we apply the **finite population correction**. The formula (from Cochran 1977) is:

$$n' = \frac{n}{1 + (n - 1/N)} \quad \textbf{Formula 7.6}$$

where n is the sample size calculated from formula 7.5; n' (read: n -prime) is the new value for the sample size; and N is the size of the total population from which n is being drawn.

Here’s an example. Suppose you are sampling the 540 resident adult men

in a Mexican village to determine how many have ever worked illegally in the United States. How many of those men do you need to interview to ensure a 95% probability sample, with a 5% confidence interval? Answer: Since we have no idea what the percentage is that we're trying to estimate, we set P and Q at .5 each in formula 7.5. Solving for n (sample size), we get:

$$n = (1.96)^2(.5)(.5)/(.05)^2 = 384.16$$

which we round up to 385. Then we apply the finite population correction:

$$n' = \frac{385}{1 + (384/540)} = 225$$

This is still a hefty percentage of the 540 people in the population, but it's a lot smaller than the 385 called for by the standard formula.

Settling for Bigger Confidence Intervals

If we were willing to settle for a 10% confidence interval, we'd need only 82 people in this example, but the trade-off would be substantial. If 65 out of 225, or 29%, reported that they had worked illegally in the United States, we would be 68% confident that from 24% to 34% really did, and 95% confident that 19% to 39% did. But if 24 out of 82 (the same 29%) reported having participated in extreme sports, we'd be 68% confident that the true figure was between 19% and 39%, and 95% confident that it was between 9% and 49%. With a spread like that, you wouldn't want to bet much on the sample statistic of 29%.

If it weren't for ethnography, this would be a major problem in taking samples from small populations—the kind we often study in anthropology. If you've been doing ethnography in a community of 1,500 people for 6 months, however, you may feel comfortable taking a confidence interval of 10% because you are personally (not statistically) confident that your intuition about the group will help you interpret the results of a small sample.

Another Catch

All of this discussion has been about estimating single parameters, whether proportions or means. You will often want to measure the interaction among several variables at once. Suppose you study a population of wealthy, middle-class, and poor people. That's three kinds of people. Now add two sexes, male and female (that makes six kinds of people) and two colors, black and white (that makes 12 kinds). If you want to know how all those independent vari-

ables combine to predict, say, average number of children desired, the sampling strategy gets more complicated.

Representative sampling is one of the trickiest parts of social research. I recommend strongly that you consult an expert in sampling if you are going to do complex tests on your data. (For fuller coverage of sampling theory and sample design, see Sudman 1976, Jaeger 1984, and Levy and Lemeshow 1999.)

8



Nonprobability Sampling and Choosing Informants

If your objective is to estimate a parameter or a proportion from a sample to a larger population, and if your research calls for the collection of data about attributes of individuals (whether those individuals are people or organizations or episodes of a sitcom), then the rule is simple: Collect data from a sufficiently large, randomly selected, unbiased sample. If you know that you *ought* to use an unbiased sample, and you have the means to *get* an unbiased sample, and you still choose to use a nonprobability sample, then expect to take a lot of flak.

There are, however, three quite different circumstances under which a nonprobability sample is exactly what is called for:

1. Nonprobability samples are always appropriate for labor-intensive, in-depth studies of a few cases. Most studies of narratives are based on fewer than 50 cases, so every case has to count. This means choosing cases on purpose, not randomly. In-depth research on sensitive topics requires nonprobability sampling. It can take months of participant observation fieldwork before you can collect narratives about topics like sexual and reproductive history or bad experiences with mental illness or use of illegal drugs.

Come to think of it, just about everything is a sensitive topic when you dig deeply enough. Sexual history is an obviously sensitive topic, but so is the management of household finances when you get into how people really allocate their resources. People love to talk about their lives, but when you get into the details of a life history, you quickly touch a lot a nerves. Really in-

depth research requires informed informants, not just responsive respondents—that is, people whom you choose on purpose, not randomly.

2. Nonprobability samples are also appropriate for large surveys when, despite our best efforts, we just can't get a probability sample. In these cases, use a nonprobability sample and *document the bias*. That's all there is to it. No need to agonize about it.
3. And, as I said at the beginning of chapter 6, when you are collecting cultural data, as contrasted with data about individuals, then expert informants, not randomly selected respondents, are what you really need. Think of the difference between asking someone "How old was your child when you first gave him an egg to eat?" versus "At what age do children here first eat eggs?" I deal with the problem of selecting cultural experts (people who are likely to really know when most mothers introduce eggs around here) in the second part of this chapter.

The major nonprobability sampling methods are: **quota sampling**, **purposive** or **(judgment) sampling**, **convenience** (or **haphazard**) **sampling**, and **chain referral** (including **snowball** and **respondent-driven**) **sampling**. Finally, **case control sampling** combines elements of probability and nonprobability sampling.

Quota Sampling

In quota sampling, you decide on the subpopulations of interest and on the proportions of those subpopulations in the final sample. If you are going to take a sample of 400 adults in a small town in Japan, you might decide that, because gender is of interest to you as an independent variable, and because women make up about half the population, then half your sample should be women and half should be men. Moreover, you decide that half of each gender quota should be older than 40 and half should be younger; and that half of each of those quotas should be self-employed and half should be salaried.

When you are all through designing your quota sample, you go out and fill the quotas. You look for, say, five self-employed women who are over 40 years of age and who earn more than 300,000 yen a month and for five salaried men who are under 40 and who earn less than 300,000 yen a month. And so on.

Tinsley et al. (2002) interviewed 437 elderly users of Lincoln Park in Chicago. They selected quota samples of about 50 men and 50 women from each of the four major ethnic groups in the area, Blacks, Whites, Hispanics, and Asian Americans. Besides gender and ethnicity, Tinsley et al. stratified on place and time. They divided the park into three zones (north, south, and middle) and three time periods (6 A.M. to 10 A.M., 11 A.M. to 3 P.M., and 4 P.M. to 8 P.M.). There were, then, nine zone-time strata in which interviewers selected

respondents. The interviewers were also told to make sure they got some weekday and some weekend users of the park.

Commercial polling companies use quota samples that are fine-tuned on the basis of decades of research (Weinberger 1973). Organizations like Gallup, Roper, Harris, and others have learned how to train interviewers not to choose respondents who are pretty much like themselves; not to select only people whom they would enjoy interviewing; not to avoid people whom they would find obnoxious or hostile; not to avoid people who are hard to contact (busy people who are hardly ever home, or people who work nights and sleep days); and not to favor people who are eager to be interviewed.

The result is quota samples that are *not unbiased* but that often do a good job of reflecting the population parameters of interest. In other words, quota sampling is an art that often approximates the results of probability sampling at less cost and less hassle than strict probability sampling.

Often, but not always. In 1948, pollsters predicted, on the basis of quota sampling, that Thomas Dewey would beat Harry Truman in the U.S. presidential election. The *Chicago Tribune* was so confident in those predictions that they printed an edition announcing Dewey's victory—while the votes were being counted that would make Truman president.

Skip to 1992. In the general election in Britain that year, four different polls published on the day of the election put the Liberal Party, on average, about 1 point ahead of the Conservative Party. All the polls were based on quota sampling. The Conservatives won by 8 points. In fact, from 1992 to 1997, political polls using quota samples in Britain systematically overestimated the support for the Liberals (Curtice and Sparrow 1997).

Quota samples are biased toward people you can find easily. This means that quota sampling is simply dangerous when it comes to making predictions about election outcomes—or estimating any population parameter, for that matter.

On the other hand, quota sampling is appropriate in the study of cultural domains. If you want to know how junior sports—Little League Baseball, Pop Warner football, Youth Soccer, junior and senior high school football—function in small communities across the United States, you'd ask people who have children playing those sports. There will be some intracultural variation, but open-ended interviews with four or five really knowledgeable people will produce the relevant cultural data—including data on the range of ideas that people have about these institutions.

Many studies of narratives are based on small samples, simply because there is so much work involved. If you are doing narrative analysis, set up a quota sampling design. First, figure out how many narratives you can collect, transcribe, and code for themes. More about all this in chapter 17 on text anal-

ysis, but narratives, like life histories, can take several interviews and many hours just to collect. Figure on 6–8 hours to transcribe each recorded hour when you start out; you’ll cut the time in half as you get better at it, assuming you have decent typing skills, and you’ll cut it in half again if you’re willing to spend the time it takes to train voice recognition software—see chapter 9. And when you get through transcribing, there’s still coding to do (more hours) and analysis and write up (lots more hours).

Suppose you think you can do 40 in-depth interviews. That means you can have up to three independent, binary variables if you want five cases in each cell, and that tells you what your sampling design will look like. Suppose you are studying, through narratives, the lived experiences of labor migrants to the United States who are back home in their community in Mexico. You decide that you want to compare the experiences of (1) people who spent time on jobs in the United States but were caught by the U.S. Border Patrol and deported with those of (2) people who managed to stay on their jobs until they’d accumulated enough money to return on their own. You also want to compare the experiences of Indians and mestizos and of men and women. That’s three binary independent variables: deported/not deported; Indian/mestizo; male/female. There are, as you can see from table 8.1, 16 cells in this

TABLE 8.1
Sampling Design for Three Dichotomous Variables

Mexicans who have worked in the United States and returned home							
Caught by the Border Patrol and deported							
Indian		Mestizo		Indian		Mestizo	
Male	Female	Male	Female	Male	Female	Male	Female
Returned home on their own							
Indian		Mestizo		Indian		Mestizo	
Male	Female	Male	Female	Male	Female	Male	Female

design. If you want to compare more variables simultaneously, you’ll have more cells on the bottom row and you’ll need more data.

Purposive or Judgment Sampling

In purposive sampling, you decide the purpose you want informants (or communities) to serve, and you go out to find some. This is somewhat like

quota sampling, except that there is no overall sampling design that tells you how many of each type of informant you need for a study. You take what you can get.

I used purposive sampling in my study of the Kalymnian (Greek) sponge-fishing industry (1987). I knew I had to interview sponge merchants, boat owners, and divers, but my first interviews taught me that I had to interview people whom I had never considered: men who used to be divers but who had quit, gone to Australia as labor migrants, and returned to their island. It was very easy to find those returned migrants: Everyone on the island either had one in their family or knew people who did.

There are many good reasons for using purposive samples. They are used widely in (1) **pilot studies**, (2) **intensive case studies**, (3) **critical case studies**, and (4) studies of **hard-to-find populations**.

1. Pilot studies. These are studies done before running a larger study. In 1999, Katherine Browne, Carla Freeman, and Zobeida Bonilla began a comparative ethnographic study of women entrepreneurs in Martinique, Barbados, and Puerto Rico—that is, in the French-, English-, and Spanish-speaking Caribbean. In a large, multisite study like this, it pays to spend time on pilot research. Each member of the team did 30 in-depth interviews with women who were engaged in a wide range of enterprises, who were of different ages, and who came from one- and two-parent homes. This helped the team develop their research instruments and provided the baseline for the larger project (Browne 2001).

And speaking of instruments, when you do surveys to test hypotheses, you want to make sure that you test all your scales with a pilot sample. More about all this in chapter 12.

2. In intensive case studies, the object is often to identify and describe a cultural phenomenon. Dickerson et al. (2000) studied the experiences of American Indian graduate nursing students and cultural barriers that might lead the students to drop out of their training. Dickerson et al. found and interviewed 11 students who were enrolled in an advanced nurse practitioner program. Samples don't get much more purposive than this, and they don't get much more appropriate, either.

Life history research and qualitative research on special populations (drug addicts, trial lawyers, shamans) rely on judgment sampling. Barroso (1997), for example, studied a purposive sample of 14 men and 6 women in the Tampa, Florida, area, all of whom had lived with AIDS for at least 3 years.

Finally, researchers don't usually pull research sites—villages, tribal encampments, hospitals, school systems—out of a hat. They rely on their judgment to find one that reflects the things they are interested in.

3. Critical case studies. Polling companies try to identify communities across the United States that have voted for the winner in the past, say, six presidential elections. Then they poll those few communities that meet the criterion.

Choosing key informants in ethnographic research is also critical-case sampling. It would be pointless to select a handful of people randomly from a population and try to turn them into trusted key informants.

4. We almost always have to rely on purposive sampling in the study of hard-to-find populations.

Think about locating and interviewing refugees from Somalia and Ethiopia living in a large American city. Many of these people experienced torture and don't exactly welcome researchers who want to ask them a lot of questions. This was the problem facing researchers in Minneapolis (see Spring et al. 2003; Jaranson et al. 2004). The study design called for a quota sample of 1,200 respondents, including 300 Oromo women, 300 Oromo men, 300 Somali women, and 300 Somali men. The study team recruited male and female interviewers from the community—people who shared ethnicity, language, and religion with the people they were trying to locate and interview. The project team sent out fliers, placed announcements in church bulletins, and made presentations at meetings of Oromo and Somali organizations. The interviewers also used their own social networks to locate potential respondents. Over 25 months, the team built trust in the community and wound up with 1,134 of the 1,200 interviews called for in the study.

Kimberly Mahaffy (1996) was interested in how lesbian Christians deal with the cognitive dissonance that comes from being rejected by mainstream Christian churches. Mahaffy sent letters to gay Christian organizations, asking them to put an ad for potential respondents in their newsletters. She sent flyers to women's bookstores and to lesbian support groups, asking for potential respondents to get in touch with her.

Eventually, Mahaffy got 163 completed questionnaires from women who fit the criteria she had established for her research, including 44 from women who self-identified as born-again or evangelical Christians. Mahaffy could not possibly have gotten an unbiased sample of lesbian Christians, but the corpus of data that she collected from her respondents had all the information she needed to answer her research questions.

Convenience or Haphazard Sampling

Convenience sampling is a glorified term for grabbing whoever will stand still long enough to answer your questions. Sometimes, convenience samples

are all that's available, and you just have to make do. Studies of the homeless, for example, are usually done with convenience samples, for obvious reasons, as are studies of people who are in intensive care units in hospitals. All samples represent *something*. The trick is to make them representative of what *you* want them to be. That's what turns a convenience sample into a purposive one.

For example, Al-Krenawi and Wiesel-Lev (1999) wanted to understand the emotions of Israeli Bedouin women who had experienced genital mutilation. They interviewed a convenience sample of 12 women who had been through the ritual and 12 women who had not but had either seen it first-hand or had heard about women in their own extended families going through it. We wouldn't put much stock in the fact that a specific *percentage* of the women reported sexual problems or relationship problems with various members of their family, but the *list* of problems is very instructive because it is the basis for more in-depth research.

If you want to estimate a parameter, then you know what you have to do: get a random, representative sample. If you want to know the percentage of adult men in a matrilineal, cross-cousin society who have actually married their biological mother's-brother's-sister (MBZ), you'll either have to count them all or take a random, unbiased sample of sufficient size to be able to make that generalization.

Key informants will tell you that the rule is broken regularly, but not by how much. A convenience sample of women who gather at the village well each day will tell you the range of options for men who don't have a biological MBZ, but not how many choose each option. And if you want to know the effect of a new road on some peasants and you only interview people who come to town on the road, you'll miss all the people who live too far off the road for it to do them any good.

Chain Referral, or Network Sampling: The Snowball and RDS Methods

Snowball and **respondent-driven sampling (RDS)** are two network sampling methods (also known, generically, as **chain referral** methods) for studying hard-to-find or hard-to-study populations. Populations can be hard to find and study for three reasons: (1) they contain very few members who are scattered over a large area (think strict vegans in rural Georgia); and/or (2) they are stigmatized and reclusive (HIV-positive people who never show up at clinics until they are sick with AIDS) or even actively hiding (intravenous drug users, for example); and/or (3) they are members of an elite group and don't care about your need for data.

In the snowball technique, you use key informants and/or documents to locate one or two people in a population. Then, you ask those people to (1) list others in the population and (2) recommend someone from the list whom you might interview. You get handed from informant to informant and the sampling frame grows with each interview. Eventually, the sampling frame becomes saturated—that is, no new names are offered.

David Griffith and his colleagues used *two* snowball samples in their study of food preferences in Moberly, Missouri. They chose an initial “seed” household in a middle-income neighborhood and asked a man in the house to name three people in town with whom he interacted on a regular basis. The first person cited by the informant lived in a lower-income neighborhood across town. That person, in turn, named other people who were in the lower-income bracket.

After a while, the researchers realized that, though they’d started with a middle-income informant who had children at home, they were getting mostly lower-income, elderly people in the snowball sample. So they started again, this time with a seed from an elite, upper-middle-income neighborhood. By the time they got through, Griffith et al. had a well-balanced sample of 30 informants with whom they did in-depth interviews (reported in Johnson 1990:78).

Thomas Weisner has been following 205 counterculture women and their families since 1974. Weisner built this sample by recruiting women in California who were in their third trimester of pregnancy. He used snowball sampling, but to ensure that participants came from all over the state and represented various kinds of families, he used no more than two referrals from any one source (Weisner 2002:277).

Snowball sampling is popular and fun to do, but in large populations it does not produce a random, representative sample. If you are dealing with a relatively small population of people who are likely to be in contact with one another, like practitioners of alternative medicine in a small town, then snowball sampling is an effective way to build an exhaustive sampling frame. Once you have an exhaustive sampling frame, you can select people at random to interview. In this case, snowball sampling is one step in a two-step process for getting a representative sample.

For large populations, however, people who are well known have a better chance of being named in a snowball procedure than are people who are less well known. And in large populations, people who have large networks name more people than do people who have small networks. For large populations, then, snowball sampling is risky because every person does not have the same chance of being included.

Douglas Heckathorn (1997) developed respondent-driven sampling for

dealing with these problems. Like snowball sampling, RDS begins with a few informants who act as seeds. The informants are paid for being interviewed and are then asked to recruit up to three members of their networks into the study. To move this process along, Heckathorn paid each of his seed informants \$10 and gave them three coupons. Anyone who came to Heckathorn to be interviewed and who had one of those coupons was paid the same \$10. (He upped the bounty to \$15 for referring a female drug injector, since they were harder to find.) Those informants, in turn, got several coupons and recruited others into the study.

There are several very important improvements to snowball sampling here. First, this method avoids the ethical problem that snowball sampling presents. The people whom an informant names may not want you even to know about their existence, much less be anxious to grant you an interview. Second, having members of a hard-to-study population do the recruiting deals with the reluctance of some people to be interviewed. And finally, Heckathorn (1997, 2002) shows that, when it's done right, the RDS method produces samples that are less biased than are traditional snowball samples. (For more on chain referral sampling, see Sudman and Kalton 1986, Martin and Dean 1993, Heckathorn and Jeffri 2001, and Salganik and Heckathorn 2004.)

Case Control Sampling

In case control sampling, you choose a purposive sample on the basis of some criterion (like having a certain illness or injury, or attempting suicide, or being homeless) and match the members of that sample with people who match the cases on many criteria, but *not* on the case criterion. This method is widely used in public health research.

For example, Beautrais et al. (1998) had a set of 129 cases of men and women under 25 who had survived a medically serious attempt at suicide in Christchurch, New Zealand. (A medically serious attempt at suicide generally involves treatment in a hospital for more than 24 hours.) Beautrais et al. recruited 153 control cases from the electoral rolls of the area, such that the mean age of the controls matched the mean age of the attempted suicide cases and the proportion of men and women in the two samples were more or less the same. In this study, both the study cases and the control cases were purposive.

Case control sampling has great potential for field research as well. Bassuk and Rosenberg (1988) wanted to know why whole families in Boston were homeless. By canvassing homeless shelters, they identified 49 female-headed, homeless families with a total of 86 children. Then they selected 81

control cases—poor, female-headed families in the same city. They made sure that the two samples—the cases and the controls—were about the same in average age of the mother (28 and 29), average age when the mother had her first child (20 and 19), and average number of children (2.4 and 2.5). Then these researchers looked for *differences* between the cases and the controls to see what might account for homelessness.

And case control is not limited to places like Christchurch and Boston. Pfeiffer et al. (2001) used the method to study why some Shona children, in Sussundenga, Mozambique, are malnourished, while others thrive. Mothers of children under 5 years of age in Sussundenga brought their children in regularly to a local clinic. Using a table like that in appendix A, Pfeiffer et al. took a random sample of 50 undernourished children (no more than one from a household) from the clinic's records. These are the **index cases**.

Next, they walked away from each of the index homes in a random direction (you just select a random number between 1 and 4) and visited every house along that path until they found the first suitable **control case**. They looked for households with: (1) one or more children under five, none of whom were malnourished; (2) more or less the same income as one of the index-case households; and (3) the same house construction (same kind of latrine, same roof material), and the same water access as one of the index-case households. The control case was the oldest child under 5 in the control household.

When they started the research, Pfeiffer et al. hypothesized that thriving children would be in households where women brought in more money (or managed more money) than women do in index-case households. The hypothesis was not supported. Instead, control mothers had about double the average education of case mothers and mothers in the control households reported more than twice the number of protein consumption days in the past month (meat, fish, poultry) as did mothers in the index households.

Sampling and Credibility

Particularly in ethnographic research, you learn in the field, as you go along, to select the units of analysis (people, court records, whatever) that will provide the information you need. This is what Russell Belk et al. (1988) did in their detailed ethnographic study of buyers and sellers at a swap meet. When you study a process, like bargaining over goods, and you're doing the research in the field, in real time (not under simulated conditions in a lab), then selecting informants who meet certain criteria is the right thing to do.

The credibility of research results comes from the power of the methods used in measurement and sampling. Good measurement is the key to internal

validity and representative sampling is the key to external validity. Well-done nonprobability sampling is actually part of good measurement. It contributes to credibility by contributing to internal validity. When someone reads a research report based on really good measurement of a nonprobability sample, they come away thinking, “Yep, I believe those conclusions about the people who were studied in that piece of research.”

That’s plenty. If you want the credibility of your conclusions to extend beyond the group of people (or countries, or organizations, or comic books) you studied, then either: (1) Repeat the study one or more times with nonprobability samples; or (2) Use a probability sample.

Choosing Informants

Across the social sciences, you’ll see references to research participants as “respondents,” or “subjects,” or “informants.” These terms tend to be used by sociologists, psychologists, and anthropologists, respectively. Respondents respond to survey questions, subjects are the subject of some experiment, and informants . . . well, informants tell you *what they think you need to know* about their culture.

There are two kinds of informants: **key informants** and **specialized informants**. Key informants are people who know a lot about their culture and are, for reasons of their own, willing to share all their knowledge with you. When you do long-term ethnography you develop close relationships with a few key informants—relationships that can last a lifetime. You don’t choose these people. They and you choose each other, over time.

Specialized informants have particular competence in some cultural domain. If you want to know the rules of Balinese cockfighting, or how many cows must be given to a Lumasaba bride’s parents, or when to genuflect in a Roman Catholic Mass, or what herb tea to give children for diarrhea, you need to talk to people who can speak knowledgeably about those things.

Key Informants

Good key informants are people whom you can talk to easily, who understand the information you need, and who are glad to give it to you or get it for you. Pelto and Pelto (1978:72) advocate training informants “to conceptualize cultural data in the frame of reference” that you, the researcher, use.

In some cases, you may want to just listen. But when you run into a really great informant, I see no reason to hold back. Teach the informant about the

analytic categories you're developing and ask whether the categories are correct. In other words, encourage the informant to become the ethnographer.

I've worked with Jesús Salinas since 1962. In 1971, I was about to write an ethnography of his culture, the *Ñähñu* of central Mexico, when he mentioned that he'd be interested in writing an ethnography himself. I dropped my project and taught him to read and write *Ñähñu*. Over the next 15 years, Salinas produced four volumes about the *Ñähñu* people—volumes that I translated and from which I learned many things that I'd never have learned had I written the ethnography myself. For example, *Ñähñu* men engage in rhyming duels, much like the “dozens” of African Americans. I wouldn't have thought to ask about those duels because I had never witnessed one (see Bernard and Salinas Pedraza 1989).

Just as Salinas has influenced my thinking about Mexican Indian life, Salinas's ethnography was heavily influenced by his association with me. We've discussed and analyzed parts of *Ñähñu* culture over the years and we've even argued over interpretation of observed facts. (More about all this in the section on native ethnography in chapter 17 on text analysis, plus a different perspective by Harry Wolcott [1999].)

Finding Key Informants

One of the most famous key informants in the ethnographic literature is Doc in William Foote Whyte's *Street Corner Society* (1981 [1943]). Whyte studied “Cornerville,” an Italian American neighborhood in a place he called “Eastern City.” (Cornerville was the North End of Boston.) Whyte asked some social workers if they knew anyone who could help Whyte with his study. One social worker told Whyte to come to her office and meet a man whom she thought could do the job. When Whyte showed up, the social worker introduced him to Doc and then left the room. Whyte nervously explained his predicament, and Doc asked him “Do you want to see the high life or the low life?” (Whyte 1989:72).

Whyte couldn't believe his luck. He told Doc he wanted to see all he could, learn as much as possible about life in the neighborhood. Doc told him:

Any nights you want to see anything, I'll take you around. I can take you to the joints—the gambling joints. I can take you around to the street corners. Just remember that you're my friend. That's all they need to know. I know these places and if I tell them you're my friend, nobody will bother you. You just tell me what you want to see, and we'll arrange it. . . . When you want some information, I'll ask for it, and you listen. When you want to find out their philosophy of life, I'll start an argument and get it for you. (ibid.)

Doc was straight up; he told Whyte to rely on him and to ask him anything, and Doc was good to his word all through Whyte's 3 years of fieldwork. Doc introduced Whyte to the boys on the corner; Doc hung out with Whyte and spoke up for Whyte when people questioned Whyte's presence. Doc was just spectacular.

Or was he? Boelen (1992) visited Cornerville 25 times between 1970 and 1989, sometimes for a few days, other times for several months. She tracked down and interviewed everyone she could find from *Street Corner Society*. Doc had died in 1967, but she interviewed his two sons in 1970 (then in their late teens and early 20s). She asked them what Doc's opinion of Whyte's book had been and reports the elder son saying: "My father considered the book untrue from the very beginning to the end, a total fantasy" (Boelen 1992:29).

Of course, Whyte (1996a, 1996b) refuted Boelen's report, but we'll never know the whole truth. Whyte certainly made mistakes, but the same can be said for all ethnographers. For some scholars, mistakes invalidate a positivist stance in ethnography. For others, it does not.

Doc may be famous, but he's not unique. He's not even rare. All successful ethnographers will tell you that they eventually came to rely on one or two key people in their fieldwork. What was rare about Doc is how quickly and easily Whyte teamed up with him. It's not easy to find informants like Doc. When Jeffrey Johnson began fieldwork in a North Carolina fishing community, he went to the local marine extension agent and asked for the agent's help. The agent, happy to oblige, told Johnson about a fisherman whom he thought could help Johnson get off on the right foot.

It turned out that the fisherman was a transplanted northerner; he had a pension from the Navy; he was an activist Republican in a thoroughly Democratic community; and he kept his fishing boat in an isolated moorage, far from the village harbor. He was, in fact, maximally different from the typical local fisherman. The agent had meant well, of course (Johnson 1990:56).

In fact, the first informants with whom you develop a working relationship in the field may be "deviant" members of their culture. Agar (1980b:86) reports that during his fieldwork in India, he was taken on by the *naik*, or headman of the village. The *naik*, it turned out, had *inherited* the role, but he was not respected in the village and did not preside over village meetings. This did not mean that the *naik* knew nothing about village affairs and customs; he was what Agar called a "solid insider," and yet somewhat of an outcast—a "marginal native," just like the ethnographer was trying to be (Freilich 1977). If you think about it, Agar said, you should wonder about the kind of person who would befriend an ethnographer.

In my own fieldwork (at sea, in Mexican villages, on Greek islands, in rural communities in the United States, and in modern American bureaucracies), I

have consistently found the best informants to be people who are cynical about their own culture. They may not be outcasts (in fact, they are always solid insiders), but they say they *feel* somewhat marginal to their culture, by virtue of their intellectualizing of and disenchantment with their culture. They are always observant, reflective, and articulate. In other words, they invariably have all the qualities that I would like to have myself.

Don't choose key ethnographic informants too quickly. Allow yourself to go awash in data for a while and play the field. When you have several prospects, check on their roles and statuses in the community. Be sure that the key informants you select don't prevent you from gaining access to other important informants (i.e., people who won't talk to you when they find out you're so-and-so's friend). Since good ethnography is, at its best, a good story, find trustworthy informants who are observant, reflective, and articulate—who know how to tell good stories—and stay with them. In the end, ethnographic fieldwork stands or falls on building mutually supportive relations with a few key people.

Informants Sometimes Lie

Don't be surprised if informants lie to you. Jeffrey Johnson, a skilled boat builder, worked in an Alaskan boatyard as part of his field study of a fishing community. At one point in his fieldwork, two other ethnographers showed up, both women, to conduct some interviews with the men in the boatyard. "The two anthropologists had no idea I was one of *them*," Johnson reports, "since I was dressed in carpenter's overalls, with all the official paraphernalia—hammer, tape measure, etc. I was sufficiently close to overhear the interview and, knowing the men being interviewed, recognized quite a few blatant lies. In fact, during the course of one interview, a captain would occasionally wink at me as he told a whopper of a lie" (personal communication).

This is not an isolated incident. A Comox Indian woman spent 2 hours narrating a text for Franz Boas. The text turned out to be nothing but a string of questions and answers. Boas didn't speak Comox well enough to know that he was being duped, but when he found out he noted it in his diary (Rohner 1969:61).

In 1938, Melville Herskovits published his massive, two-volume work on the ancient West African kingdom of Dahomey (today Benin). According to Herskovits, there was an annual census and the data from these efforts were used in administering the state. The counting involved the delivery of sacks of pebbles from around the kingdom to the palace at Abomey, with each pebble representing a person. Roger Sandall (1999) has shown that the informant who

told Herskovits about this elaborate accounting system may have made it all up.

This sort of thing can happen to anyone who does participant observation ethnography, but some cultures are more tolerant of lying than are others. Nachman (1984) found that the most articulate informants among the Nissan of New Guinea were great truth tellers and accomplished liars at the same time. Among the Nissan, says Nachman, people expect big men to give speeches and to “manipulate others and to create socially acceptable meanings,” even if that means telling outright lies (*ibid.*:552).

Selecting Culturally Specialized Informants

The search for formal and systematic ways to select focused ethnographic informants—people who can help you learn about particular areas of a culture—has been going on for a very long time. In 1957, Marc-Adelard Tremblay was involved in a Cornell University survey research project on poverty in Nova Scotia. He wanted to use ethnographic informants to help the team’s researchers design a useful questionnaire, so he made a list of some roles in the community he was studying—things like sawmill owners, doctors, farmers, bankers—and chose informants who could talk to him knowledgeably about things in their area of expertise. Tremblay had no external test to tell him whether the informants he selected were, in fact, the most competent in their areas of expertise, but he felt that on-the-spot clues made the selection of informants valid (Tremblay 1957).

Michael Robbins and his colleagues studied acculturation and modernization among the Baganda of Uganda, using a more formal method to select informants who might be competent on this topic (Robbins et al. 1969). First, they ran a survey of households in a rural sector, asking about things that would indicate respondents’ exposure to Western culture. Then they used the results of the survey to select appropriate informants.

Robbins et al. had 80 variables in the survey that had something to do with acculturation and they ran a factor analysis to find out which variables package together. We’ll look a bit more at factor analysis in chapter 21. For now, think of factor analysis as a way to reduce those 80 variables to just a handful of underlying variables around which individual variables cluster. It turned out that 14 of the original 80 variables clustered together in one factor. Among those original variables were: being under 40 years of age, drinking European beer, speaking and reading English, having a Western job, and living in a house that has concrete floors and walls.

Robbins et al. called this cluster the “acculturation factor.” They chose

informants who had high scores on this factor and interviewed them about acculturation. Robbins et al. reversed Tremblay’s method. Tremblay used key informants to help him build a survey instrument; Robbins et al. used a survey to find key informants.

In any given domain of culture, some people are more competent than others. In our culture, some people know a lot about the history of baseball; some people can name the actors in every sitcom since the beginning of television in the 1940s. Some people are experts on medicinal plants; others are experts on cars and trucks. John Poggie (1972) did an early study of informant competence. He selected one informant in each of seven Mexican communities. The communities ranged in size from 350 to 3,000 inhabitants. The informants were village or town presidents, or judges, or (in the case of agricultural communities) the local commissioners of communal land. Poggie asked these informants questions about life in the communities, and he compared the answers with data from a high-quality social survey.

For example, Poggie asked the seven informants: “How many men in this town are workers in Ciudad Industrial?” (Ciudad Industrial is a fictitious name of a city that attracted many labor migrants from the communities that Poggie studied.) In his survey, Poggie asked respondents if they had ever worked in Ciudad Industrial. The correlation between the answers given by Poggie’s expert informants and the data obtained from the survey was .90.

Poggie also asked: “What percentage of the houses here are made of adobe?” This time, the correlation between the informants and the survey was only .71. Table 8.2 shows the seven questions Poggie asked, and how well his informants did when their answers were compared to the survey.

TABLE 8.2
Agreement between Informants and Survey Data in Seven Villages

<i>Questions asked of informants</i>	<i>Correlation with questionnaire data</i>
Number of men from this town who are workers in Ciudad Industrial	0.90
Percentage of houses made of adobe	0.71
Percentage of households that have radios	0.52
Percentage of people who eat eggs regularly	0.33
Percentage of people who would like to live in Ciudad Industrial	0.23
Percentage of people who eat bread daily	0.14
Percentage of people who sleep in beds	0.05

SOURCE: J. J. Poggie, “Toward Quality Control in Key Informant Data,” *Human Organization*, Vol. 31, pp. 26–29, 1972. Reprinted with permission of the Society for Applied Anthropology.

Overall, informants produced answers most like those in the survey when they were asked to respond to questions about things that are publicly observable. The survey data are not necessarily more *accurate* than the informants' data. But as the questions require informants to talk about things inside people's homes (such as what percentage of people eat eggs), or about what people think (what percentage of people would *like* to live in Ciudad Industrial), informants' answers look less and less like those of the survey.

Poggie concluded that "There is little reason to believe that trust and rapport would improve the reliability and precision concerning what percentage sleep in beds, who would like to live in the new industrial city, or what percentage eat bread daily" (ibid.:29).

The Cultural Consensus Model

The idea that people can be more or less competent in various areas of their culture has led to formal tests of new methods for selecting focused ethnographic informants. James Boster (1985, 1986) walked 58 Aguaruna Jívaro women (in Peru) through a garden that had 61 varieties of manioc. He asked the women *waji mama aita?* ("What kind of manioc is this?"), and calculated the likelihood that all possible pairs of women agreed on the name of a plant. Since Boster had planted the garden himself, he *knew* the true identification of each plant. Sure enough, the more that women agreed on the identification of a plant, the more likely they were to know what the plant actually was. In other words, as cultural consensus increased, so did cultural competence.

This makes a lot of sense. Suppose you give a test about the rules of baseball to two groups of people: a group of rabid baseball fans and another group (Americans, Canadians, Mexicans, Dominicans, etc.) who never watch the game. You'd expect that: (1) The serious baseball fans will agree more among themselves about the answers to your test questions than will the nonfans; and (2) The serious fans will get the answers right more often than the nonfans. These outcomes are expected because of the relation between cultural consensus and cultural competence.

Boster's experiment and the hypothetical baseball experiment are pretty much like any test you might take in a class. The instructor makes up both the test and an answer key with the (supposedly) correct answers. Your job is to match your answers with those on the answer key.

But what if there were no answer key? That's what happens when we ask people to tell us the uses of various plants, or to list the sacred sites in a village, or to rate the social status of others in a community. We are not asking people for their opinions, attitudes, beliefs, or values. We ask informants to

rate the social status of others in their community because we want to *know* the social status of all those people. The problem is, we don't have an answer key to tell whether informants are accurate in their reporting of information.

Romney et al. (1986) developed a formal method, called the **cultural consensus model**, to test informant competence *without having an answer key*. The theory behind the technique makes three assumptions:

1. Informants share a common culture and there is a culturally correct answer to any question you ask them. The culturally correct answer might be incorrect from an outsider's perspective (as often happens when we compare folk knowledge about illnesses to biomedical knowledge). Any variation you find among informants is the result of *individual* differences in their knowledge, not the result of being members of subcultures.
2. Informants give their answers to your test questions independently of one another.
3. All the questions in your test come from the same cultural domain—that is, things that can be listed, like kinds of animals or hand tools, or things you can do on a weekend. (We'll take up cultural domain analysis in chapter 11.) A test that asks about kinship and Australian-rules football would be a poor test. People can be competent in one domain and incompetent in another. The cultural consensus method must be used *only* for identifying people who are knowledgeable about a particular domain.

To use the consensus technique, simply give a sample of informants a test that asks them to make some judgments about a list of items in a cultural domain. You can use true-false and yes-no questions. An example of a true-false question in fieldwork might be: "You can get [pneumonia] [diarrhea] [*susto*] from [being overweight] [tired] [scared] [in the room with a sick person]." Some other typical true-false questions might be: "You can't get AIDS from touching the body of someone who died from it," or "A field goal is worth 7 points."

You can also use multiple-choice questions or even open-ended, fill-in-the-blank questions. (See appendix F for information about *Anthropac*, a set of programs that includes modules for handling cultural consensus data.)

For the test to reliably distinguish cultural competence among informants, it's best to have about 40 test items and about 40 informants. As an example, table 8.3 shows the answers of four informants to a 40-question true-false test about "general knowledge" for Americans (things like who starred in some classic movies). The 1s are items to which a student answered "true" (or "yes"), and the 0s are items to which a student answered "false" (or "no").

Table 8.4 shows the *number* of matches between informants, the *proportion of matches* (the number of matches divided by the number of items in the

TABLE 8.4
Matches, Proportions of Matches, Proportions of Corrected Matches, and
Competency Scores for the Data in Table 8.3

<i>Informant</i>	<i>Number of matches (matrix I)</i>				<i>Proportion of matches (matrix II)</i>				<i>Proportion of corrected matches (matrix III)</i>				<i>Competency score for student</i>	
	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>		
	1	—	27	24	22	—	0.675	0.625	0.550	—	0.35	0.25		0.10
2	27	—	34	21	0.675	—	0.850	0.525	0.35	—	0.70	0.05	2	0.61
3	25	34	—	23	0.625	0.850	—	0.575	0.25	0.70	—	0.15	3	0.61
4	22	21	23	—	0.550	0.525	0.575	—	0.10	0.05	0.15	—	4	0.32

SOURCE: "Culture as Consensus: A Theory of Culture and Informant Accuracy" by A. K. Romney et al., 1986, *American Anthropologist* 88:324. Reproduced by permission of the American Anthropological Association. Not for further reproduction.

test), and the proportion of matches *corrected for guessing*. This correction is necessary because anyone can guess the answers to any true-false test item half the time. *Anthropac* has a built-in error-correction routine for consensus analysis.

The three matrices in table 8.4 are called **similarity matrices** because the entries in each matrix give some direct estimate of how similar any pair of informants is (see chapters 11, 16, and 21 for more on similarity matrices). Look at Matrix I, the one called “number of matches.” Informants 1 and 2 have 27 matches. If you look along the first two rows of table 8.3 and count, you’ll see that on 27 out of 40 test questions, informants 1 and 2 answered the same. When informant 1 said “false” (0), then informant 2 said “false” (0), and when informant 1 said “true” (1), then informant 2 said “true” (1).

Now look at Matrix II, “proportion of matches.” This shows that informants 1 and 2 were 67.5% similar, because $27/40 = .675$. Finally, look at Matrix III, “proportion of corrected matches.” After correcting for the possibility that some of the similarity in Matrix II between informants is due to the fact that they guessed the same answers when they didn’t really know the answers, we see that informants 1 and 2 are .35 alike, while informants 2 and 3 are .70 alike. Informants 2 and 3 are twice as similar to one another as informants 1 and 2 are to one another.

Look down the last column of Matrix III. Informant 4 is not like any other informant. That is, informant 4’s answers to the 40 questions were practically idiosyncratic compared to the answers that other informants gave.

We can use this information to compute a competency score for each informant. To do this, run a factor analysis on the matrix of corrected matches. (*Anthropac* does all this automatically. You don’t need to understand factor analysis to read the rest of this section. For an introduction to factor analysis, see chapter 21.) If the three conditions I’ve listed for the model have been met, then the first factor in the solution should be at least three times the size of the second factor. If it is, then this means that: (1) The first factor is *knowledge* about the domain (because agreement equals knowledge under conditions of the model); and (2) The individual factor scores are a measure of knowledge for each person who takes the test.

At the far right of table 8.4, we see that informants 2 and 3 have the highest factor scores (.61). They are also the students who got the highest of the four scores in the general knowledge test. You can use the consensus test on any group of informants, for any cultural domain. Triad tests, paired comparisons, ratings, and rankings all produce data that can be subjected to consensus analysis, as do true-false tests and multiple choice tests.

I want to stress that if you are doing general descriptive ethnography, and you’re looking for all-around good informants, the cultural consensus method

is *not* a substitute for the time-honored way that ethnographers have always chosen key informants: luck, intuition, and hard work by both parties to achieve a working relationship based on trust. The cultural consensus method, though, is truly useful for finding highly competent people who can talk about well-defined areas of cultural knowledge. (For a detailed explanation of the math behind consensus analysis, see Weller 2004 and <http://www.analytictech.com/borgatti/consensu.htm>.)

Testing the Cultural Consensus Model

The cultural consensus model makes a lot of sense, but it may be a bit of a stretch to imagine that you can find the answer key to a test under certain conditions. You can test this. Get the results from any multiple-choice test in any class that has at least 40 students and run the consensus analysis available in *Anthropac*.

Correlate the first factor score for each student against the score that each student actually got on the test. If the exam was a good test of student's knowledge (that is, if the set of exam questions represents the cultural domain), you'll get a correlation of over .90. What that means is that the students who have the highest first-factor scores (knowledge scores) will mirror the professor's answer key for at least 90% of the items.

If you can retrieve an etically correct answer key, then you can apply the model (cautiously, of course, always cautiously) to tests of emic data, like people's ideas about who hangs out with whom in an organization or what people think are good ways to cure a cold, avoid getting AIDS, take care of a baby, etc.

The cultural consensus model is an important contribution to social science methods. It means that, under the conditions of the model (informants share a common culture and there is a cultural answer to each question; informants answer test questions independently of one another; the questions in the test come from a single cultural domain), you can build the answer key to a test from the matrix of agreements among informants. (For more about the consensus model, see Weller 2004. For more examples of consensus analysis, see Caulkins 2001, de Munck et al. 2002, Furlow 2003, Swora 2003, Harvey and Bird 2004, Jaskyte and Dressler 2004, and Miller et al. 2004.)

Selecting Domain-Specific Informants

Weller and Romney (1988), two of the developers of the cultural consensus model, have determined the number of informants you need to produce valid

and reliable data about particular cultural domains, given that the three conditions of the model are more-or-less met. (I say “more-or-less” because the model is very robust, which means that it produces very similar answers even when its conditions are more-or-less, not perfectly, met.)

Table 8.5 shows those numbers: Just 10 informants, with an average competence of .7 have a 99% probability of answering each question on a true-false

TABLE 8.5

Minimal Number of Informants Needed to Classify a Desired Proportion of Questions with a Specified Confidence Level for Different Levels of Cultural Competence

Proportion of questions	Average level of cultural competence				
	.5	.6	.7	.8	.9
<i>.95 confidence level</i>					
0.80	9	7	4	4	4
0.85	11	7	4	4	4
0.90	13	9	6	4	4
0.95	17	11	6	6	4
0.99	29	19	10	8	4
<i>.99 confidence level</i>					
0.80	15	10	5	4	4
0.85	15	10	7	5	4
0.90	21	12	7	5	4
0.95	23	14	9	7	4
0.99	>30	20	13	8	6

SOURCE: S. C. Weller and A. K. Romney, *Systematic Data Collection*, p. 77. © 1988. Reprinted by permission of Sage Publications.

test correctly, with a confidence level of .95. Only 13 informants, with a relatively low average competence of .5 are needed if you want a 90% probability of answering each question on a test correctly, with a confidence level of .95.

Weller and Romney also (1988) showed that you can use the simple Spearman-Brown Prophecy formula, available in many general statistical packages, as a proxy for the full consensus method (the one that involves doing a factor analysis on the informant-by-informant agreement matrix, and so on) when you have interval level data. Table 8.6 shows the results: If you interview 10 informants whose responses correlate .49, then the aggregate of their answers are likely to correlate .95 with the true answers.

Adam Kiš (2005) studied funerals in a village in Malawi. Funerals used to be events that brought everyone in the community together, but AIDS has changed that. There are now so many funerals that people have to decide

TABLE 8.6
 Agreement among Individuals and Estimated Validity of Aggregating
 Their Responses for Different Samples

Agreement	Validity				
	0.80	0.85	0.90	0.95	0.99
0.16	10	14	22	49	257
0.25	5	8	13	28	148
0.36	3	5	8	17	87
0.49	2	3	4	10	51

SOURCE: S. C. Weller and A. K. Romney, *Systematic Data Collection*, p. 77. © 1988. Reprinted by permission of Sage Publications.

which ones to attend. Kiš developed a cultural domain test of “reasons to attend a funeral” and administered it to 30 informants. The results showed that there was, indeed, a consensus about this domain of funeral culture, so Kiš focused on the most knowledgeable informants, as determined by the consensus analysis, for his in-depth, ethnographic interviews.

Caution: If you use consensus analysis to find knowledgeable informants, watch out for the **shaman effect**. People who have very specialized knowledge about some field may be very different in their knowledge profile from people in the mainstream—that is, shamans. In fact, it is to the advantage of shamans everywhere, whether their knowledge is about curing illness or making money on the stock market, to protect that knowledge by keeping it maximally different from mainstream knowledge. The bottom line: Use consensus analysis to find the highly knowledgeable informants, but never pass up the chance to interview a shaman.

Lots more about consensus analysis in chapter 11 on cultural domain analysis.

Paying Informants

Finally, there’s the issue of whether to pay informants, and if so, how much? If you are studying people who are worth millions of dollars, paying them is inappropriate. You can’t possibly pay them enough to compensate them financially for their time. Better to make a donation to a charity that they support. This will vary from case to case, but the general rule, for me at least, is that if you want to interview people, they should be paid at the local rate for their time. And speaking of interviews. . . .

9



Interviewing: Unstructured and Semistructured

The Big Picture

The concept of “interviewing” covers a lot of ground, from totally unstructured interactions, through semistructured situations, to highly formal interactions with respondents. Interviewing is done on the phone, in person, by mail—even by computer. This chapter is about unstructured and semistructured face-to-face interviewing, including the management of focus groups.

Unstructured interviewing goes on all the time and just about anywhere—in homes, walking along a road, weeding a millet field, hanging out in bars, or waiting for a bus. **Semistructured**, or **in-depth interviewing** is a scheduled activity. A semistructured interview is open ended, but follows a general script and covers a list of topics.

There is a vast literature on how to conduct effective interviews: how to gain rapport, how to get people to open up, how to introduce an interview, and how to end one. You can’t learn to interview by reading about it, but after you read this chapter, and practice some of the techniques described, you should be well on your way to becoming an effective interviewer. You should also have a pretty good idea of how much more there is to learn, and be on your way to exploring the literature.

Interview Control

There is a continuum of interview situations based on the **amount of control** we try to exercise over people’s responses (Dohrenwend and Richardson

1965; Gorden 1975; Spradley 1979). These *different types* of interviews produce *different types* of data that are useful for *different types* of research projects and that appeal to *different types* of researchers. For convenience, I divide the continuum of interviews into four large chunks.

1. Informal Interviewing

At one end there is **informal interviewing**, characterized by a total lack of structure or control. The researcher just tries to remember conversations heard during the course of a day in the field. This requires constant jotting and daily sessions in which you sit at a computer, typing away, unburdening your memory, and developing field notes. Informal interviewing is the method of choice at the beginning of participant observation fieldwork, when you're settling in. It is also used throughout ethnographic fieldwork to build greater rapport and to uncover new topics of interest that might have been overlooked.

When it comes to interviewing, never mistake the adjective "informal" for "lightweight." This is hard, hard work. You have to remember a lot; you have to duck into private corners a lot (so you can jot things down); and you have to use a lot of deception (to keep people from knowing that you're really at work, studying them). Informal interviewing can get pretty tiring.

Still, in some kinds of research, informal interviewing is all you've got. Mark Connolly (1990) studied *gamines*, or street children, in Guatemala City, Guatemala, and Bogotá, Colombia. These children live, eat, and sleep on the street. Hanging out and talking informally with these children was an appropriate way to do this research. Informal ethnography can also be combined with more structured methods, when circumstances allow it. In fact, Rachel Baker (1996a, 1996b) was able to collect anthropometric data on street children in Kathmandu, Nepal, while doing informal ethnography.

2. Unstructured Interviewing

Next comes **unstructured interviewing**, one of the two types covered in this chapter. There is nothing at all informal about unstructured interviewing, and nothing deceptive, either. You sit down with another person and hold an interview. Period. Both of you know what you're doing, and there is no shared feeling that you're just engaged in pleasant chitchat.

Unstructured interviews are based on a clear plan that you keep constantly in mind, but are also characterized by a minimum of control over the people's responses. The idea is to get people to open up and let them express themselves in their own terms, and at their own pace. A lot of what is called **ethnographic interviewing** is unstructured. Unstructured interviewing is used in

situations where you have lots and lots of time—like when you are doing long-term fieldwork and can interview people on many separate occasions.

3. *Semistructured Interviewing*

In situations where you won't get more than one chance to interview someone, **semistructured interviewing** is best. It has much of the freewheeling quality of unstructured interviewing, and requires all the same skills, but semistructured interviewing is based on the use of an **interview guide**. This is a written list of questions and topics that need to be covered in a particular order.

This is the kind of interview that most people write about—the kind done in professional surveys. The interviewer maintains discretion to follow leads, but the interview guide is a set of clear instructions—instructions like this one: “Probe to see if informants (men and women alike) who have daughters have different values about dowry and about premarital sex than do people who have only sons.”

Formal, written guides are an absolute must if you are sending out several interviewers to collect data. But even if you do all the interviewing on a project yourself, you should build a guide and follow it if you want reliable, comparable qualitative data.

Semistructured interviewing works very well in projects where you are dealing with high-level bureaucrats and elite members of a community—people who are accustomed to efficient use of their time. It demonstrates that you are fully in control of what you want from an interview but leaves both you and your respondent free to follow new leads. It shows that you are prepared and competent but that you are not trying to exercise excessive control.

4. *Structured Interviewing*

Finally, in fully **structured interviews**, people are asked to respond to as nearly identical a set of stimuli as possible. One variety of structured interviews involves use of an **interview schedule**—an explicit set of instructions to interviewers who administer questionnaires orally. Instructions might read: “If the informant says that she or he has at least one daughter over 10 years of age, then ask questions 26b and 26c. Otherwise, go on to question 27.”

Questionnaires are one kind of structured interview. Other structured interviewing techniques include pile sorting, frame elicitation, triad sorting, and tasks that require informants to rate or rank order a list of things. I'll deal with structured interviews in chapter 10.

Unstructured Interviewing

Unstructured interviewing is truly versatile. It is used equally by scholars who identify with the hermeneutic tradition and by those who identify with the positivist tradition. It is used in studies that require only textual data and in studies that require both textual and numerical data. Ethnographers may use it to develop formal guides for semistructured interviews, or to learn what questions to include, in the native language, on a highly structured questionnaire (see Werner and Schoepfle [1987] for a good discussion of this). I say that ethnographers *may* use unstructured interviewing in developing structured interview schedules because unstructured interviewing also stands on its own.

When you want to know about the **lived experience** of fellow human beings—what it’s like to survive hand-to-hand combat, how you get through each day when you have a child dying of leukemia, how it feels to make it across the border into Texas from Mexico only to be deported 24 hours later—you just can’t beat unstructured interviewing.

Unstructured interviewing is excellent for building initial rapport with people, before moving to more formal interviews, and it’s perfect for talking to informants who would not tolerate a more formal interview. The personal rapport you build with close informants in long-term fieldwork can make highly structured interviewing—and even semistructured interviewing—feel somehow unnatural. In fact, really structured interviewing can get in the way of your ability to communicate freely with key informants.

But not always. Some people want very much to talk about their lives, but they really don’t like the unstructured interview format. I once asked a fisherman in Greece if I could have a few minutes of his time to discuss the economics of small-scale fishing. I was about 5 minutes into the interview, treading lightly—you know, trying not to get too quickly into his finances, even though that’s exactly what I wanted to know about—when he interrupted me: “Why don’t you just get to the point?” he asked. “You want to know how I decide where to fish, and whether I use a share system or a wage system to split the profits, and how I find buyers for my catch, and things like that, right?” He had heard from other fishermen that these were some of the topics I was interviewing people about. No unstructured interviews for him; he was a busy man and wanted to get right to it.

A Case Study of Unstructured Interviewing

Once you learn the art of probing (which I’ll discuss in a bit), unstructured interviewing can be used for studying sensitive issues, like sexuality, racial or ethnic prejudice, or hot political topics. I find it particularly useful in studying

conflict. In 1972–1973, I went to sea on two different oceanographic research vessels (Bernard and Killworth 1973, 1974). In both cases, there was an almost palpable tension between the scientific personnel and the crew of the ship. Through both informal and unstructured interviewing on land between cruises, I was able to establish that the conflict was predictable and regular. Let me give you an idea of how complex the situation was.

In 1972–1973, it cost \$5,000 a day to run a major research vessel, not including the cost of the science. (That would be about \$25,000 today.) The way oceanography works, at least in the United States, the chief scientist on a research cruise has to pay for both ship time and for the cost of any experiments he or she wants to run. To do this, oceanographers compete for grants from institutions like the U.S. Office of Naval Research, NASA, and the National Science Foundation.

The spending of so much money is validated by publishing significant results in prominent journals. It's a tough, competitive game and one that leads scientists to use every minute of their ship time. As one set of scientists comes ashore after a month at sea, the next set is on the dock waiting to set up their experiments and haul anchor.

The crew, consequently, might only get 24 or 48 hours shore leave between voyages. That can cause some pretty serious resentment by ships' crews against scientists. And that can lead to disaster. I found many documented instances of sabotage of expensive research by crew members who were, as one of them said, "sick and tired of being treated like goddamn bus drivers." In one incident, involving a British research vessel, a freezer filled with Antarctic shrimp, representing 2 years of data collection, went overboard during the night. In another, the crew and scientists from a U.S. Navy oceanographic research ship got into a brawl while in port (*Science* 1972:1346).

The structural problem I uncovered began at the top. Scientists whom I interviewed felt they had the right to take the vessels wherever they wanted to go, within prudence and reason, in search of answers to questions they had set up in their proposals. The captains of the ships believed (correctly) that *they* had the last word on maneuvering their ships at sea. Scientists, said the captains, sometimes went beyond prudence and reason in what they demanded of the vessels.

For example, a scientist might ask the captain to take a ship out of port in dangerous weather because ship time is so precious. This conflict between crew and scientists has been known—and pretty much ignored—since Charles Darwin sailed with HMS *Beagle* and it will certainly play a role in the productivity of long-term space station operations.

Unraveling this conflict at sea required participant observation and unstructured (as well as informal) interviewing with many people. No other strategy

for data collection would have worked. At sea, people live for weeks, or even months, in close quarters, and there is a common need to maintain good relations for the organization to function well.

It would have been inappropriate for me to have used highly structured interviews about the source of tension between the crew and the scientists. Better to steer the interviews around to the issue of interest and to let informants teach me what I needed to know. In the end, no analysis was better than that offered by one engine room mechanic who told me, "These scientist types are so damn hungry for data, they'd run the ship aground looking for interesting rocks if we let them."

Getting Started

There are some important steps to take when you start interviewing someone for the first time. First of all, assure people of anonymity and confidentiality. Explain that you simply want to know what *they* think, and what *their* observations are. If you are interviewing someone whom you have come to know over a period of time, explain why you think their opinions and observations on a particular topic are important. If you are interviewing someone chosen from a random sample, and whom you are unlikely to see again, explain how they were chosen and why it is important that you have their cooperation to maintain representativeness.

If people say that they really don't know enough to be part of your study, assure them that their participation is crucial and that you are truly interested in what they have to say (and you'd better mean it, or you'll never pull it off). Tell everyone you interview that you are trying to learn from *them*. Encourage them to interrupt you during the interview with anything they think is important. And always ask for permission to *record* personal interviews *and to take notes*. This is vital. If you can't take notes, then, in most cases, the value of an interview plummets. (See below, on using a tape recorder and taking notes.)

Keep in mind that people who are being interviewed know that you are shopping for information. There is no point in trying to hide this. If you are open and honest about your intentions, and if you are genuinely interested in what people have to say, many people will help you.

This is not always true, though. When Colin Turnbull went out to study the Ik in Uganda, he found a group of people who had apparently lost interest in life and in exchanging human kindnesses. The Ik had been brutalized, decimated, and left by the government to fend for themselves on a barren reservation. They weren't impressed with the fact that Turnbull wanted to study their culture. In fact, they weren't much interested in anything Turnbull was up to and were anything but friendly (Turnbull 1972).

Letting the Informant or Respondent Lead

If you can carry on “unthreatening, self-controlled, supportive, polite, and cordial interaction in everyday life,” then interviewing will come easy to you, and informants will feel comfortable responding to your questions (Lofland 1976:90). But no matter how supportive you are as a person, an interview is never really like a casual, unthreatening conversation in everyday life. In casual conversations, people take more or less balanced turns (Spradley 1979), and there is no feeling that somehow the discussion has to stay on track or follow some theme (see also Merton et al. 1956; Hyman and Cobb 1975). In unstructured interviewing, you keep the conversation focused on a topic, while giving the respondent room to define the content of the discussion.

The rule is: Get people on to a topic of interest and get out of the way. Let the informant provide information that he or she thinks is important.

During my research on the Kalymnian sponge fishermen in Greece, I spent a lot of time at Procopis Kambouris’s *taverna*. (A Greek *taverna* is a particular kind of restaurant.) Procopis’s was a favorite of the sponge fishermen. Procopis was a superb cook, he made his own wine every year from grapes that he selected himself, and he was as good a teller of sea stories as he was a listener to those of his clientele. At Procopis’s *taverna*, I was able to collect the work histories of sponge fishermen—when they’d begun their careers, the training they’d gotten, the jobs they’d held, and so on. The atmosphere was relaxed (plenty of retsina wine and good things to eat), and conversation was easy.

As a participant observer, I developed a sense of camaraderie with the regulars, and we exchanged sea stories with a lot of flourish. Still, no one at Procopis’s ever made the mistake of thinking that I was there just for the camaraderie. They knew that I was writing about their lives and that I had lots of questions to ask. They also knew immediately when I switched from the role of participant observer to that of ethnographic interviewer.

One night, I slipped into just such an interview/conversation with Savas Ergas. He was 64 years old at the time and was planning to make one last 6-month voyage as a sponge diver during the coming season in 1965. I began to interview Savas on his work history at about 7:30 in the evening, and we closed Procopis’s place at about 3 in the morning. During the course of the evening, several other men joined and left the group at various times, as they would on any night of conversation at Procopis’s. Savas had lots of stories to tell (he was a living legend and he played well to a crowd), and we had to continue the interview a few days later, over several more liters of retsina.

At one point on that second night, Savas told me (almost offhandedly) that he had spent more than a year of his life walking the bottom of the Mediterra-

nean. I asked him how he knew this, and he challenged me to document it. Savas had decided that there was something important that I needed to know and he maneuvered the interview around to make sure I learned it.

This led to about 3 hours of painstaking work. We counted the number of seasons he'd been to sea over a 46-year career (he remembered that he hadn't worked at all during 1943 because of "something to do with the war"). We figured conservatively the number of days he'd spent at sea, the average number of dives per trip, and the average depth and time per dive. We joked about the tendency of divers to exaggerate their exploits and about how fragile human memory is when it comes to this kind of detail.

It was difficult to stay on the subject, because Savas was such a good raconteur and a perceptive analyst of Kalymnian life. The interview meandered off on interesting tangents, but after a while, either Savas or I would steer it back to the issue at hand. In the end, discounting heavily for both exaggeration and faulty recall, we reckoned that he'd spent at least 10,000 hours—about a year and a fourth, counting each day as a full 24 hours—under water and had walked the distance between Alexandria and Tunis at least three times.

The exact numbers really didn't matter. What did matter was that Savas Ergas had a really good sense of what *he* thought I needed to know about the life of a sponge diver. It was I, the interviewer, who defined the focus of the interview; but it was Savas, the respondent, who determined the content. And was I ever glad he did.

Probing

The key to successful interviewing is learning how to probe effectively—that is, to stimulate a respondent to produce more information, without injecting yourself so much into the interaction that you only get a reflection of yourself in the data. Suppose you ask, "Have you ever been away from the village to work?" and the informant says, "Yes." The next question (the probe) is: "Like where?" Suppose the answer is, "Oh, several different places." The correct response is not, "Pachuca? Querétaro? Mexico City?" but, "Like where? Could you name some of the places where you've gone to get work?"

There are many kinds of probes that you can use in an interview. (In what follows, I draw on the important work by Kluckhohn [1945], Merton et al. [1956], Kahn and Cannell [1957], Whyte [1960, 1984], Dohrenwend and Richardson [1965], Gorden [1975], Hyman and Cobb [1975], Warwick and Lininger [1975], Reed and Stimson [1985], and on my own experience and that of my students.)

The Silent Probe

The most difficult technique to learn is the **silent probe**, which consists of just remaining quiet and waiting for an informant to continue. The silence may be accompanied by a nod or by a mumbled “uh-huh” as you focus on your note pad. The silent probe sometimes produces more information than does direct questioning. At least at the beginning of an interview, informants look to you for guidance as to whether or not they’re on the right track. They want to know whether they’re “giving you what you want.” Most of the time, especially in unstructured interviews, you want the informant to define the relevant information.

Some informants are more glib than others and require very little prodding to keep up the flow of information. Others are more reflective and take their time. Inexperienced interviewers tend to jump in with verbal probes as soon as an informant goes silent. Meanwhile, the informant may be just reflecting, gathering thoughts, and preparing to say something important. You can kill those moments (and there are a lot of them) with your interruptions.

Glibness can be a matter of *cultural*, not just personal style. Gordon Streib reports that he had to adjust his own interviewing style radically when he left New York City to study the Navajo in the 1950s (Streib 1952). Streib, a New Yorker himself, had done studies based on semistructured interviews with subway workers in New York. Those workers maintained a fast, hard-driving pace during the interviews—a pace with which Streib, as a member of the culture, was comfortable.

But that style was entirely inappropriate with the Navajo, who were uniformly more reflective than the subway workers (Streib, personal communication). In other words, the silent probe is sometimes not a “probe” at all; being quiet and waiting for an informant to continue may simply be appropriate cultural behavior.

On the other hand, the silent probe is a high-risk technique, which is why beginners avoid it. If an informant is genuinely at the end of a thought and you don’t provide further guidance, your silence can become awkward. You may even lose your credibility as an interviewer. The silent probe takes practice to use effectively. But it’s worth the effort.

The Echo Probe

Another kind of probe consists of simply repeating the last thing someone has said, and asking them to continue. This **echo probe** is particularly useful when an informant is describing a process, or an event. “I see. The goat’s throat is cut and the blood is drained into a pan for cooking with the meat.

Then what happens?” This probe is neutral and doesn’t redirect the interview. It shows that you understand what’s been said so far and encourages the informant to continue with the narrative. If you use the echo probe too often, though, you’ll hear an exasperated informant asking, “Why do you keep repeating what I just said?”

The Uh-huh Probe

You can encourage an informant to continue with a narrative by just making affirmative comments, like “Uh-huh,” or “Yes, I see,” or “Right, uh-huh,” and so on. Matarazzo (1964) showed how powerful this **neutral probe** can be. He did a series of identical, semistructured, 45-minute interviews with a group of informants. He broke each interview into three 15-minute chunks. During the second chunk, the interviewer was told to make affirmative noises, like “uh-huh,” whenever the informant was speaking. Informant responses during those chunks were about a third longer than during the first and third periods.

The Tell-Me-More Probe

This may be the most common form of probe among experienced interviewers. Respondents give you an answer, and you probe for more by saying: “Could you tell me more about that?” Other variations include “Why exactly do you say that?” and “Why exactly do you feel that way?” You have to be careful about using stock probes like these. As Converse and Schuman point out (1974:50), if you get into a rut and repeat these probes like a robot, don’t be surprised to hear someone finishing up a nice long discourse by saying, “Yeah, yeah, and why *exactly* do I feel like that?” (From personal experience, I can guarantee that the mortification factor only allows this sort of thing to happen once. The memory of the experience lasts a lifetime.)

The Long Question Probe

Another way to induce longer and more continuous responses is by making your questions longer. Instead of asking, “How do you plant a home garden?” ask, “What are all the things you have to do to actually get a home garden going?” When I interviewed sponge divers on Kalymnos, instead of asking them, “What is it like to make a dive into very deep water?” I said, “Tell me about diving into really deep water. What do you do to get ready and how do you descend and ascend? What’s it like down there?”

Later in the interview or on another occasion, I would home in on special

topics. But to break the ice and get the interview flowing, there is nothing quite as useful as what Spradley (1979) called the **grand tour question**.

This does not mean that asking longer questions or using neutral probes necessarily produces *better* responses. They do, however, produce *more* responses, and, in general, more is better. Furthermore, the more you can keep an informant talking, the more you can express interest in what they are saying and the more you build rapport. This is especially important in the first interview you do with someone whose trust you want to build (see *ibid.*:80). There is still a lot to be learned about how various kinds of probes affect what informants tell us.

Threatening questions—those asking for sensitive information—should be short but preceded by a long, rambling run-up: “We’re interested in the various things that people do these days in order to keep from getting diseases when they have sex. Some people do different kinds of things, and some people do nothing special. Do you ever use condoms?” If the respondent says, “Yes,” or “No,” or “Sometimes,” *then* you can launch that series of questions about why, why not, when, with whom, and so on. The wording of sensitive questions should be supportive and nonjudgmental. (See below for more on threatening questions.)

Probing by Leading

After all this, you may be cautious about being really directive in an interview. Don’t be. Many researchers caution against “leading” an informant. Lofland (1976), for example, warns against questions like, “Don’t you think that? . . .” and suggests asking, “What do you think about? . . .” He is, of course, correct. On the other hand, any question an interviewer asks leads an informant. You might as well learn to do it well.

Consider this leading question that I asked a Nāhñu Indian: “Right. I understand. The compadre is *supposed* to pay for the music for the baptism fiesta. But what happens if the compadre doesn’t have the money? Who pays then?” This kind of question can stop the flow of an informant’s narrative stone dead. It can also produce more information than the informant would otherwise have provided. At the time, I thought the informant was being overly “normative.” That is, I thought he was stating an ideal behavioral custom (having a compadre pay for the music at a fiesta) as if it were never violated.

It turned out that all he was doing was relying on his own cultural competence—“abbreviating,” as Spradley (1979:79) called it. The informant took for granted that the anthropologist knew the “obvious” answer: If the compadre didn’t have enough money, well, then there might not be any music.

My interruption reminded the informant that I just wasn’t up to his level of

cultural competence; I needed him to be more explicit. He went on to explain other things that he considered obvious but that I would not have even known to ask about. Someone who has committed himself to pay for the music at a fiesta might borrow money from *another* compadre to fulfill the obligation. In that case, he wouldn't tell the person who was throwing the fiesta. That might make the host feel bad, like he was forcing his compadre to go into debt.

In this interview, in fact, the informant eventually became irritated with me because I asked so many things that he considered obvious. He wanted to abbreviate a lot and to provide a more general summary; I wanted details. I backed off and asked a different informant for the details. I have since learned to start some probes with "This may seem obvious, but. . . ."

Directive probes (leading questions) may be based on what an informant has just finished saying, or may be based on something an informant told you an hour ago, or a week ago. As you progress in long-term research, you come to have a much greater appreciation for what you really want from an interview. It is perfectly legitimate to use the information you've already collected to focus your subsequent interviews.

This leads researchers from informal to unstructured to semistructured interviews and even to completely structured interviews like questionnaires. When you feel as though you have learned something important about a group and its culture, the next step to test that knowledge—to see if it is idiosyncratic to a particular informant or subgroup in the culture or if it can be reproduced in many informants.

Baiting: The Phased-Assertion Probe

A particularly effective probing technique is called **phased assertion** (Kirk and Miller 1986), or **baiting** (Agar 1996:142). This is when you act like you already know something in order to get people to open up.

I used this technique in a study of how Nāhñu Indian parents felt about their children learning to read and write Nāhñu. Bilingual (Spanish-Indian) education in Mexico is a politically sensitive issue (Heath 1972), and when I started asking about it, a lot of people were reluctant to talk freely.

In the course of informal interviewing, I learned from a schoolteacher in one village that some fathers had come to complain about the teacher trying to get the children to read and write Nāhñu. The fathers, it seems, were afraid that studying Nāhñu would get in the way of their children becoming fluent in Spanish. Once I heard this story, I began to drop hints that I knew the reason parents were against children learning to read and write Nāhñu. As I did this, the parents opened up and confirmed what I'd found out.

Every journalist (and gossip monger) knows this technique well. As you

learn a piece of a puzzle from one informant, you use it with the next informant to get more information, and so on. The more you seem to know, the more comfortable people feel about talking to you and the less people feel they are actually divulging anything. *They* are not the ones who are giving away the “secrets” of the group.

Phased assertion also prompts some informants to jump in and correct you if they think you know a little, but that you’ve “got it all wrong.” In some cases, I’ve purposely made wrong assertions to provoke a correcting response.

Verbal Respondents

Some people try to tell you *too much*. They are the kind of people who just love to have an audience. You ask them one little question and off they go on one tangent after another, until you become exasperated. Converse and Schuman (1974:46) recommend “gentle inattention”—putting down your pen, looking away, leafing through your papers. Nigel King (1994:23) recommends saying something like: “That’s very interesting. Could we go back to what you were saying earlier about. . . .”

You may, however, have to be a bit more obvious. New interviewers, in particular, may be reluctant to cut off informants, afraid that doing so is poor interviewing technique. In fact, as William Foote Whyte notes, informants who want to talk your ear off are probably used to being interrupted. It’s the only way their friends get a word in edgewise. But you need to learn how to cut people off without rancor. “Don’t interrupt *accidentally* . . .,” Whyte said, “learn to interrupt *gracefully*” (1960:353, emphasis his). Each situation is somewhat different; you learn as you go in this business.

Nonverbal Respondents

One of the really tough things you run into is someone telling you “I don’t know” in answer to lots of questions. In qualitative research projects, where you choose respondents precisely because you think they know something of interest, the “don’t know” refrain can be especially frustrating. Converse and Schuman (1974:49) distinguish four kinds of don’t-know response: (1) I don’t know (and frankly I don’t care); (2) I don’t know (and it’s none of your business); (3) I don’t know (actually, I do know, but you wouldn’t be interested in what I have to say about that); and (4) I don’t know (and I wish you’d change the subject because this line of questioning makes me really uncomfortable). There is also the “(I wish I could help you but) I really don’t know.”

Sometimes you can get beyond this, sometimes you can’t. You have to face the fact that not everyone who volunteers to be interviewed is a good respon-

dent. If you probe those people for information when they say, “I don’t know,” you tempt them to make something up just to satisfy you, as Sanchez and Morchio (1992) found. Sometimes, you just have to take the “don’t know” for an answer and cut your losses by going on to someone else.

The Ethics of Probing

Are these tricks of the trade ethical? I think they are, but using them creates some responsibilities to your respondents.

First, there is no ethical imperative in social research more important than seeing to it that you do not harm innocent people who have provided you with information in good faith. The problem, of course, is that not all respondents are innocents. Some people commit wartime atrocities. Some practice infanticide. Some are HIV-positive and, out of bitterness, are purposely infecting others. Do you protect them all?

Are any of these examples more troublesome to you than others? These are not extreme cases, thrown in here to prepare you for the worst, “just in case.” They are the sorts of ethical dilemmas that field researchers confront all the time.

Second, the better you get at making people “open up,” the more responsible you become that they don’t later suffer some emotional distress for having done so. Informants who divulge *too* quickly what they believe to be secret information can later come to have real regrets and even loss of self-esteem. They may suffer anxiety over how much they can trust you to protect them in the community.

It is sometimes better to stop an informant from divulging privileged information in the first or second interview and to wait until both of you have built a mutually trusting relationship. If you sense that an informant is uncomfortable with having spoken too quickly about a sensitive topic, end the interview with light conversation and reassurances about your discretion. Soon after, look up the informant and engage in light conversation again, with no probing or other interviewing techniques involved. This will also provide reassurance of trust.

Remember: The first ethical decision you make in research is whether to collect certain kinds of information at all. Once that decision is made, *you* are responsible for what is done with that information, and *you* must protect people from becoming emotionally burdened for having talked to you.

Learning to Interview

It’s impossible to eliminate reactivity and subjectivity in interviewing, but like any other craft, you get better and better at interviewing the more you

practice. It helps a lot to practice in front of others and to have an experienced interviewer monitor and criticize your performance. Even without such help, however, you can improve your interviewing technique just by paying careful attention to what you're doing. Harry Wolcott (1995) offers excellent advice on this score: Pay as much attention to your own words as you do to the words of your respondents (p. 102).

Wolcott also advises: Keep interviews focused on a few big issues (ibid.:112). More good advice from one of the most accomplished ethnographers around. Here's a guaranteed way to wreck rapport and ruin an interview: An informant asks you, "Why do you ask? What does that have to do with what we're talking about?" You tell her: "Well, it just seemed like an interesting question—you know, something I thought might be useful somehow down the road in the analysis."

Here you are, asking people to give you their time and tell you about their lives and you're treating that time with little respect. If you can't imagine giving a satisfactory answer to the question: "Why did you ask *that*?" then leave *that* out.

Do *not* use your friends as practice informants. You cannot learn to interview with friends because there are role expectations that get in the way. Just when you're really rolling, and getting into probing deeply on some topic that you both know about, they are likely to laugh at you or tell you to knock it off.

Practice interviews should *not* be just for practice. They should be done on topics you're really interested in and with people who are likely to know a lot about those topics. Every interview you do should be conducted as professionally as possible and should produce useful data (with plenty of notes that you can code, file, and cross-file).

The Importance of Language

Most anthropologists (and an increasing number of sociologists and social psychologists) do research outside their own country. If you are planning to go abroad for research, find people from the culture you are going to study and interview them on some topic of interest. If you are going to Turkey to study women's roles, then find Turkish students at your university and interview them on some related topic.

It is often possible to hire the spouses of foreign students for these kinds of "practice" interviews. I put "practice" in quotes to emphasize again that these interviews should produce real data of real interest to you. If you are studying

a language that you'll need for research, these practice interviews will help you sharpen your skills at interviewing in that language.

Even if you are going off to the interior of the Amazon, this doesn't let you off the hook. It is unlikely that you'll find native speakers of Yanomami on your campus, but you cannot use this as an excuse to wait until you're out in the field to learn general interviewing skills. Interviewing skills are honed by practice. Among the most constructive things you can do in preparing for field research is to practice conducting unstructured and semistructured interviewing. Learn to interview in Portuguese or Spanish (depending on whether the Yanomami you are going to visit live in the Brazilian or Venezuelan Amazon) before heading for the field and you'll be way ahead.

Pacing the Study

Two of the biggest problems faced by researchers who rely heavily on semistructured interviews are boredom and fatigue. Even small projects may require 30–40 interviews to generate sufficient data to be worthwhile. Most field researchers collect their own interview data, and asking the same questions over and over again can get pretty old. Gorden (1975) studied 30 interviewers who worked for 12 days doing about two tape-recorded interviews per day. Each interview was from 1 to 2 hours long.

The first interview on each day, over all interviewers, averaged about 30 pages of transcription. The second averaged only 25 pages. Furthermore, the first interviews, on average, got shorter and shorter during the 12-day period of the study. In other words, on any given day, boredom made the second interview shorter, and over the 12 days, boredom (and possibly fatigue) took its toll on the first interviews of each day.

Even anthropologists who spend a year in the field may have focused bouts of interviewing on a particular topic. Plan each project, or subproject, in advance and calculate the number of interviews you are going to get. Pace yourself. Spread the project out if possible, and don't try to bring in all your interview data in the shortest possible time—unless you're studying reactions to a hot issue, in which case, spreading things out can create a serious history confound (see chapter 4).

Here's the tradeoff: The longer a project takes, the less likely that the first interviews and the last interviews will be valid indicators of the same things. In long-term, participant observation fieldwork (6 months to a year), I recommend going back to your early informants and interviewing them a second time. See whether their observations and attitudes have changed, and if so, why.

Presentation of Self

How should you present yourself in an interview? As a friend? As a professional? As someone who is sympathetic or as someone who is nonjudgmental? It depends on the nature of the project. When the object is to collect comparable data across respondents, then it makes no difference whether you're collecting words or numbers—cordial-but-nonjudgmental is the way to go.

That's sometimes tough to do. You're interviewing someone on a project about what people can do to help the environment, and your respondent says: "All those eco-Nazis want is to make room for more owls. They don't give a damn about real people's jobs." (Yes, that happened on one of my projects.) That's when you find out whether you can probe without injecting your feelings into the interview. Professional interviewers (the folks who collect the data for the General Social Survey, for example) learn to maintain their equilibrium and move on (see Converse and Schuman 1974).

Some situations are so painful, however, that it's impossible to maintain a neutral facade. Gene Shelley interviewed 72 people in Atlanta, Georgia, who were HIV-positive (Shelley et al. 1995). Here's a typical comment by one of Shelly's informants: "I have a lot of trouble watching all my friends die. Sometimes my whole body shuts down inside. I don't want to know people who are going to die. Some of my friends, there are three or four people a week in the obits. We all watch the obits."

How would *you* respond? Do you say: "Uh-huh. Tell me more about that"? Do you let silence take over and force the respondent to go on? Do you say something sympathetic? Shelley reports that she treated each interview as a unique situation and responded as her intuition told her to respond—sometimes more clinically, sometimes less, depending on her judgment of what the respondent needed her to say. Good advice.

On Just Being Yourself

In 1964, when we were working on the island of Kalymnos, my wife Carole would take our 2-month-old baby for daily walks in a carriage. Older women would peek into the baby carriage and make disapproving noises when they saw our daughter sleeping on her stomach. Then they would reach into the carriage and turn the baby over, explaining forcefully that the baby would get the evil eye if we continued to let her sleep on her stomach.

Carole had read the latest edition of *The Commonsense Book of Baby and Child Care* (the classic baby book by Dr. Benjamin Spock). We carried two copies of the book with us—in case one fell out of a boat or something—and Carole was convinced by Dr. Spock's writings that babies who sleep on their

backs risk choking on their own mucous or vomit. Since then, of course, medical opinion—and all the baby books that young parents read nowadays—have flip-flopped about this issue several times. At the time, though, not wanting to offend anyone, Carole listened politely and tried to act nonjudgmental.

One day, enough was enough. Carole told off a woman who intervened and that was that. From then on, women were more eager to discuss child-rearing practices in general, and the more we challenged them, the more they challenged us. There was no rancor involved, and we learned a lot more than if Carole had just kept on listening politely and had said nothing. This was informal interviewing in the context of long-term participant observation. So, if we had offended anyone, there would have been time and opportunity to make amends—or at least come to an understanding about cultural differences.

Little Things Mean a Lot

Little things are important in interviewing, so pay attention to them. How you dress and where you hold an interview, for example, tell your respondent a lot about you and what you expect. The “interviewing dress code” is: Use common sense. Proper dress depends on the venue. Showing up with a backpack or an attaché case, wearing jeans or a business suit—these are choices that should be pretty easy to make, once you’ve made the commitment to accommodate your dress to different circumstances.

Same goes for venue. I’ve held interviews in bars, in business offices, in government offices, on ferry boats, on beaches, in homes. . . . I can’t give you a rule for selecting the single *right* place for an interview, since there may be several right places. But some places are just plain wrong for certain interviews. Here again, common sense goes a long way.

Using a Voice Recorder

Don’t rely on your memory in interviewing; use a voice recorder in all structured and semistructured interviews, except where people specifically ask you not to. Recorded interviews are a permanent archive of primary information that can be passed on to other researchers. (Remember, I’m talking here about formal interviews, not the hanging-out, informal interviews that are part of ethnographic research. More on *that* in chapter 17.)

If you sense some reluctance about the recorder, leave it on the table and don’t turn it on right away. Start the interview with chitchat and when things get warmed up, say something like “This is really interesting. I don’t want to trust my memory on something as important as this; do you mind if I record

it?” Charles Kadushin (personal communication) hands people a microphone with a shut-off switch. Rarely, he says, do respondents actually use the switch, but giving people control over the interview shows that you take them very seriously.

Sometimes you’ll be recording an interview and things will be going along just fine and you’ll sense that a respondent is backing off from some sensitive topic. Just reach over to the recorder and ask the respondent if she or he would like you to turn it off. Harry Wolcott (1995:114) recommends leaving the recorder on, if possible, when the formal part of an interview ends. Even though you’ve finished, Wolcott points out, your respondent may have more to say.

Recording Equipment: Machines, Media, and Batteries

The array of recording devices available today is impressive but, as you make your choices of equipment to take to the field, remember: These are tools and only tools. Don’t get caught up by the “gee whiz” factor. If it does what you want it to do, no technology is obsolete.

There are three choices: cassette tape, minidisk (also known as MiniDisc, or MD format), and digital. They all have their pluses and minuses, though I suspect that this is the last edition of this book in which I’ll be talking about tape.

Digital has a lot going for it. Good digital recorders start at around \$75 (street price) and hold 10–15 hours of voice recording with 32mb of flash memory. When the memory is full, you upload the contents to a computer (through a USB port, for example) and then burn a CD to store your interviews offline. If you have an Apple iPod®, and if you don’t need all the disk space for music, you can turn the machine into a digital audio recorder with a plug-in microphone (see appendix F). A gigabyte of disk space holds about 400 hours of voice recordings, so a 20-gigabyte iPod has plenty of room for both music and interviews.

But caution: (1) Use the right technology, or it will take as long to upload digital audio to your computer, so you can transcribe it, as it takes to record it in the first place. (2) Okay, you have the money to hire a transcriptionist. Be sure that he or she can work from digital files. Transcribing from voice to text is traditionally done with a transcribing machine (more on them in a minute), and those machines are mostly for cassettes and microcassettes. You can make a cassette from digital audio, but it’s very time consuming. (3) If you are in an isolated field site and don’t have reliable power, digital audio can be risky. Imagine filling your digital recorder and needing to upload before you can start another interview and then . . . the power goes out, or your portable gen-

erator goes down. You'll wish then you'd stuck with a good quality, battery-operated cassette or minidisk recorder.

If you have reasonably reliable power in the field, and if you don't need a hard, paper transcription of your field notes or your interviews, then digital recording has another huge advantage: Many software packages for managing text let you code, on the fly, as you listen to digitally recorded text. In other words, they let you tag a digital recording of *voice* with digital codes, just as if you were doing it on a document on your computer. Digital audio has several minor advantages as well. Unlike tape, you can make copies of it, byte for byte, without losing any fidelity; you can post copies to the Internet to share with others; and you can insert actual sound snippets into lectures or papers presented at conferences.

The big advantage of cassettes and minidisks is that they are separate, hard media. Minidisks and microcassettes, however, are not available everywhere the way standard cassette tapes are, so if you opt for these media and you're going to the Andes, bring plenty of them with you. Many professionals still prefer top-of-the-line cassette recorders for field research, though these machines are quite expensive, compared to the alternatives available. Highly rated field machines (like the Sony Professional Walkman Minidisk) were selling for \$250–\$400 in 2005. This is professional equipment—the sort you'd want for linguistic fieldwork (when you're straining to hear every phoneme) or for high-quality recording of music. If you are not investing in professional-level equipment, there are many very good field tape recorders that cost less than \$200.

In fact, for simple recording of interviews, especially in a language you understand well, you can get away with a good, basic cassette machine for under \$50, or a microcassette machine for under \$100. But buy two of them. When you skimp on equipment costs, and don't have a spare, this almost guarantees that you'll need one at the most inconvenient moment.

Use a good, separate microphone (\$20–\$50). Some people like wearing a lavalier microphone—the kind you clip to a person's lapel or shirt collar—but many people find them intrusive. I've always preferred omnidirectional microphones (good ones cost a bit more), because they pick up voices from anywhere in a room. Sometimes, people get rolling on a topic and they want to get up and pace the room as they talk. Want to kill a really great interview? Tell somebody who's on a roll to please sit down and speak directly into the mike. Good microphones come with stands that keep the head from resting on any surface, like a table. Surfaces pick up and introduce background noise into any recording. If you don't have a really good stand for the mike, you can make one easily with some rubbery foam (the kind they use in making mattresses).

No matter what you spend on a tape or minidisk recorder, never, ever skimp on the quality of tapes or minidisks. Use only cassettes that are put together with screws so you can open them up and fix the tape when (inevitably) they jam or tangle. And don't use 120-minute tapes. Transcribing involves listening, stopping, and rewinding—often hundreds of times per tape. Thin tape (the kind that runs for 2 hours or more) just won't stand up to that kind of use.

Bruce Jackson (1987:145), a very experienced fieldworker in folklore, recommends taking brand new tapes to a studio and getting them bulk erased before recording on them for the first time. This cuts down the magnetic field noise on the new tape. Jackson also recommends running each tape through your machine three or four times on fast forward and fast reverse. All tapes stretch a bit, even the best of them, and this will get the stretch out of the way.

Test your tape recorder before every interview. And do the testing at home. There's only one thing worse than a recorder that doesn't run at all. It's one that runs but doesn't record. Then your informant is sure to say at the end of the interview: "Let's run that back and see how it came out!" (Yes, that happened to me. But only once. And it needn't happen to anyone who reads this.)

Good tape recorders have battery indicators. Want another foolproof way to kill an exciting interview? Ask the informant to "please hold that thought" while you change batteries. When batteries get slightly low, throw them out. Edward Ives (1995) recommends doing all recording on batteries. That guarantees that, no matter what kind of flaky or spiky current you run into, *your* recordings will always be made at exactly the same speed.

Particularly if you are working in places that have unstable current, you'll want to rely on batteries to ensure recording fidelity. Just make sure that you start out with fresh batteries for each interview. (You can save a lot of battery life by using house current for all playback, fast forward, and rewind operations—reserving the batteries *only* for recording.) If you prefer household current for recording, then carry along a couple of long extension cords so you have a choice of where to set up for the interview.

Good tape recorders come with voice activation (VA). When you're in VA mode, the recorder only turns on if there is noise to record. During long pauses (while an informant is thinking, for example), the recorder shuts off, saving tape. Holly Williams, however (personal communication), recommends not using the VA mode. It doesn't save much tape and she finds that the long breaks without any sound make transcribing tapes much easier. You don't have to shut the machine off and turn it on as many times while you're typing.

Transcribers

It takes 6–8 hours to transcribe 1 hour of tape, depending on how closely you transcribe (getting all the "uhs" and "ers" and throat clearings, or just

capturing the main elements of speech), how clear the tape is, and how proficient you are in the language and in typing. Invest in a transcription machine. Don't even try to transcribe taped interviews without one of those machines unless you are conducting an experiment to see how long it takes to get frustrated with transcribing. These machines cost around \$250 to \$300. You use a foot pedal to start and stop the machine, to back up and to fast forward, and even to slow down the tape so you can listen carefully to a phrase or a word. A transcription machine and a good set of earphones will save you many hours of work because you can keep both hands on your keyboard all the time.

It isn't always necessary to fully transcribe interviews. If you are using life histories to describe how families in some community deal with prolonged absence of fathers, then you *must* have full transcriptions to work with. And you can't study cultural *themes*, either, without full transcriptions. But if you want to know how many informants said they had actually used oral rehydration therapy to treat their children's diarrhea, you may be able to get away with only partial transcription. You may even be as well off using an interview guide and taking notes. (More about transcribing machines in appendix F.)

Whether you do full transcriptions or just take notes during interviews, always try to record your interviews. You may need to go back and fill in details in your notes.

Voice Recognition Software

Voice recognition software (VRS) has come of age. You listen to an interview through a set of headphones and repeat the words—both your questions and your informant's responses—out loud, in your own voice. The software listens to your voice and types out the words across the screen. You go over each sentence to correct mistakes (tell it that the word “bloat” should be “float” for instance) and to format the text (tell it where to put punctuation and paragraph breaks). The process is slow at first, but the software learns over time to recognize inflections in your voice, and it makes fewer and fewer mistakes as weeks go by. It also learns all the special vocabulary you throw at it. The built-in vocabularies of current VRS systems are enormous—something like 300,000 words—but, though they may be ready to recognize polygamy, for example, you'll have to teach it polygyny or fraternal polyandry. And, of course, you'll have to train it to recognize words from the language of your field site. If you say, “Juanita sold eight *huipiles* at the market this week,” you'll have to spell out “Juanita” and “*huipiles*” so the software can add these words to its vocabulary.

As the software gets trained, the process moves up to 95%–98% accuracy at about 100 to 120 word per minute. With a 2%–5% error rate, you still have

to go over every line of your work to correct it, but the total time for transcribing interviews can be reduced by half or more. The two most widely used products are ViaVoice™, from IBM, and ScanSoft's Dragon Naturally Speaking™ (see appendix F).

Recording Is Not a Substitute for Taking Notes

Finally, never substitute recording for note taking. A lot of very bad things can happen to tape or disks or flash memory, and if you haven't got backup notes, you're out of luck. Don't wait until you get home to take notes, either. Take notes during the interview *about* the interview. Did the informant seem nervous or evasive? Were there a lot of interruptions? What were the physical surroundings like? How much probing did you have to do? Take notes on the contents of the interview, even though you get every word on tape.

A few informants, of course, will let you use a recorder but will balk at your taking notes. Don't assume, however, that informants will be offended if you take notes. Ask them. Most of the time, all you do by avoiding note taking is lose a lot of data. Informants are under no illusions about what you're doing. You're interviewing them. You might as well take notes and get people used to it, if you can.

Focus Groups and Group Interviews

Focus groups are recruited to discuss a particular topic—anything from people's feelings about brands of beer to their experience in toilet training their children. Not all group interviews, however, are *focus group* interviews. Sometimes, you just find yourself in an interview situation with a lot of people. You're interviewing someone and other people just come up and insert themselves into the conversation. This happens spontaneously all the time in long-term fieldwork in small communities, where people all know one another. If you insist on privacy, you might find yourself with no interview at all. Better to take advantage of the situation and just let the information flow. Be sure to take notes, of course, on who's there, who's dominant, who's just listening, and so on, in any group interview.

Rachel Baker (1996a, 1996b) studied homeless boys in Kathmandu. When she interviewed boys in temples or junkyards, others might come by and be welcomed into the conversation-interview situation.

Focus groups are quite different. The method derives from work by Paul Lazarsfeld and Robert Merton in 1941 at Columbia University's Office of

Radio Research. A group of people listened to a recorded radio program that was supposed to raise public morale prior to America's entry into World War II. The listeners were told to push a red button whenever they heard something that made them react negatively and to push a green button when they heard something that made them react positively. The reactions were recorded automatically by a primitive polygraph-like apparatus. When the program was over, an interviewer talked to the group of listeners to find out why they had felt positively or negatively about each message they'd reacted to (Merton 1987).

The commercial potential of Lazarsfeld and Merton's pioneering work was immediately clear. The method of real-time recording of people's reactions, combined with focused interviewing of a group, is today a mainstay in advertising research. MCI, the long-distance phone company, used focus groups to develop their initial advertising when they were just starting out. They found that customers didn't blame AT&T for the high cost of their long-distance phone bills; they blamed themselves for talking too long on long-distance calls. MCI came out with the advertising slogan: "You're not talking too much, just spending too much." The rest, as they say, is history (Krueger 1994:33).

Whole companies now specialize in focus group research, and there are manuals on how to recruit participants and how to conduct a focus group session (Stewart and Shamdasani 1990; Krueger 1994; Vaughn et al. 1996; Morgan 1997; Morgan and Krueger 1998).

Why Are Focus Groups So Popular?

The focus group method was a commercial success from the 1950s on, but it lay dormant in academic circles for more than 20 years. This is probably because the method is virtually devoid of statistics. Since the late 1970s, however, interest among social researchers of all kinds has boomed as researchers have come to understand the benefits of combining qualitative and quantitative methods.

Focus groups do not replace surveys, but rather complement them. You can convene a focus group to discuss questions for a survey. Do the questions seem arrogant to respondents? Appropriate? Naive? A focus group can discuss the wording of a particular question or offer advice on how the whole questionnaire comes off to respondents. And you can convene a focus group to help interpret the results of a survey. But focus groups are not just adjuncts to surveys. They are widely used to find out *why* people feel as they do about something or the steps that people go through in making decisions.

Three Cases of Focus Groups

Knodel et al. (1984), for example, used focus groups to study the fertility transition in Thailand. They held separate group sessions for married men under 35 and married women under 30 who wanted three or fewer children. They also held separate sessions for men and women over 50 who had at least five children. This gave them four separate groups. In all cases, the participants had no more than an elementary school education.

Knodel et al. repeated this four-group design in six parts of Thailand to cover the religious and ethnic diversity of the country. The focus of each group discussion was on the number of children people wanted and why.

Thailand has recently undergone fertility transition, and the focus group study illuminated the reasons for the transition. “Time and again,” these researchers report, “when participants were asked why the younger generation wants smaller families than the older generation had, they responded that nowadays everything is expensive” (ibid.:302).

People also said that all children, girls as well as boys, needed education to get the jobs that would pay for the more expensive, monetized lifestyle to which people were becoming accustomed. It is, certainly, easier to pay for the education of fewer children. These consistent responses are what you’d expect in a society undergoing fertility transition.

Ruth Wilson et al. (1993) used focus groups in their study of acute respiratory illness (ARI) in Swaziland. They interviewed 33 individual mothers, 13 traditional healers, and 17 health care providers. They also ran 33 focus groups, 16 male groups and 17 female groups. The groups had from 4 to 15 participants, with an average of 7.

Each individual respondent and each group was presented with two hypothetical cases. Wilson et al. asked their respondents to diagnose each case and to suggest treatments. Here are the cases:

Case 1. A mother has a 1-year-old baby girl with the following signs: coughing, fever, sore throat, running or blocked nose, and red or teary eyes. When you ask the mother, she tells you that the child can breast-feed well but is not actively playing.

Case 2. A 10-month-old baby was brought to a health center with the following signs: rapid/difficult breathing, chest indrawing, fever for one day, sunken eyes, coughing for three days. The mother tells you that the child does not have diarrhea but has a poor appetite.

Many useful comparisons were possible with the data from this study. For example, mothers attributed the illness in Case 2 mostly to the weather, hered-

ity, or the child's home environment. The male focus groups diagnosed the child in Case 2 as having asthma, fever, indigestion, malnutrition, or worms.

Wilson et al. (1993) acknowledge that a large number of individual interviews make it easier to estimate the degree of error in a set of interviews. However, they conclude that the focus groups provided valid data on the terminology and practices related to ARI in Swaziland. Wilson and her coworkers did, after all, have 240 respondents in their focus groups; they had data from in-depth interviews of all categories of persons involved in treating children's ARI; and they had plenty of participant observation in Swaziland to back them up.

Paul Nkwi (1996), an anthropologist at the University of Yaounde, Cameroon, studied people's perceptions of family planning in his country. He and his team worked in four communities, using participant observation, in-depth interviews, a questionnaire, and focus groups. In each community, the team conducted nine focus groups on community development concerns, causes of resistance to family planning, cultural and economic factors that can be used to promote family planning, community problems with health and family planning services, how services could be improved to meet the needs of communities, and how much (if at all) people would pay for improved health care services.

The focus groups, conducted in the local language of each community, lasted from 1.5 to 2 hours and were conducted in the homes of influential men of the communities. This helped ensure that the discussions would produce useful information. The groups were stratified by age and sex. One group was exclusively young men 12–19 years of age; another group was exclusively young women of that age. Then there were male and female groups 20–35, 36–49, and 50 and over. Finally, Nkwi and his team did a single focus group with mixed ages and sexes in each community.

The focus groups were taped and transcribed for analysis. It turned out that the information from the focus groups duplicated much of the information gathered by the other methods used in the study. Nkwi's study shows clearly the value of using several data-gathering methods in one study. When several methods produce the same results, you can be a lot more secure in the validity of the findings. Nkwi's study also shows the potential for focus group interviewing in assessing public policy issues (Paul Nkwi, personal communication).

Note two very important things about all three of these cases: (1) They weren't based on a *focus group* but on a *series of groups*. Each of the groups was chosen to represent a subgroup in a factorial design, just as we saw with experiments in chapter 4 and with survey sampling in chapter 6. (2) Each of

the groups was homogeneous with respect to certain independent variables—again, just as we saw with respect to experimental and sampling design.

The principle of **factorial design** is an essential part of focus group methodology. The study by Knodel et al. (1984) on page 234 is an example of factorial design: two age groups and two genders, for a total of four groups (men under 35 and women under 30, who wanted three or fewer children, and men over 50 and women over 50, who had at least five children), repeated in six venues across Thailand, for a total of 24 groups.

Are Focus Groups Valid?

Ward et al. (1991) compared focus group and survey data from three studies of voluntary sterilization (tubal ligation or vasectomy) in Guatemala, Honduras, and Zaire. Ward et al. report that, “Overall, for 28% of the variables the results were similar” in the focus group and survey data. “For 42% the results were similar but focus groups provided additional detail; for 17% the results were similar, but the survey provided more detail. And in only 12% of the variables were the results dissimilar” (p. 273).

In the Guatemala study, 97% of the women surveyed reported no regrets with their decision to have a tubal ligation. The “vast majority” of women in the focus groups also reported no regrets. This was counted as a “similar result.” Ten percent of the women surveyed reported having had a tubal ligation for health reasons. In the focus groups, too, just a few women reported health factors in their decision to have the operation, but they provided more detail and context, citing such things as complications from previous pregnancies.

This is an example of where the focus group and survey provide similar results, but where the focus group offers more detail. Data from the focus groups and the survey confirm that women heard about the operation from similar sources, but the survey shows that 40% of the women heard about it from a sterilized woman, 26% heard about it from a health professional, and so on. Here, the survey provides more detail, though both methods produce similar conclusions.

In general, though, focus groups—like participant observation, in-depth interviews, and other systematic qualitative methods—should be used for the collection of data about content and process and should not be relied on for collecting data about personal attributes or for estimating population parameters of personal attributes. The belief that a woman has or does not have a right to an abortion is a personal attribute, like gender, age, annual income, or religion. If you want to estimate the proportion of people in a population who

believe that a woman has a right to an abortion, then focus groups are not the method of choice.

A proportion is a number, and if you want a good number—a valid one, a useful one—then you need a method that produces exactly that. A survey, based on a representative sample, is the method of choice here. But if you want information about content—about *why* people think a woman should or should not have the right to an abortion—then that’s just the sort of thing a focus group can illuminate.

Focus Group Size, Composition, Number

Focus groups typically have 6–12 members, plus a moderator. Seven or eight people is a popular size. If a group is too small, it can be dominated by one or two loudmouths; if it gets beyond 10 or 12, it gets tough to manage. However, smaller groups are better when you’re trying to get really in-depth discussions going about sensitive issues (Morgan 1997). Of course, this assumes that the group is run by a skilled moderator who knows how to get people to open up and how keep them opened up.

The participants in a focus group should be more or less homogeneous and, in general, should not know one another. Richard Krueger, a very experienced focus group moderator, says that “familiarity tends to inhibit disclosure” (1994:18). It’s easy to open up more when you get into a discussion with people whom you are unlikely ever to see again (sort of like what happens on long air flights).

Obviously, what “homogeneous” means depends on what you’re trying to learn. If you want to know why a smaller percentage of middle-class African American women over 40 get mammograms than do their white counterparts, then you need a group of middle-class African American women who are over 40.

Running a Focus Group

The group moderator gets people talking about whatever issue is under discussion. Leading a focus group requires the combined skills of an ethnographer, a survey researcher, and a therapist. You have to watch out for people who want to show off and close them down without coming on too strongly. You have to watch out for shy people and draw them out, without being intimidating.

Tips on how to do all this, and a lot more, are in *The Focus Group Kit*, a series of six how-to books (Morgan and Krueger 1998). Don’t even think about getting into focus group management without going through this kit.

In a focus group about sensitive issues like abortion or drug use, the leader works at getting the group to gel and getting members to feel that they are part of an understanding cohort of people. If the group is run by an accomplished leader, one or more members will eventually feel comfortable about divulging sensitive information about themselves. Once the ice is broken, others will feel less threatened and will join in. Moderators should not be known to the members of a focus group, and in particular, focus group members should not be employees of a moderator. Hierarchy is not conducive to openness.

In running a focus group, remember that people will disclose more in groups that are supportive and nonjudgmental. Tell people that there are no right or wrong answers to the questions you will ask and emphasize that you've invited people who are similar in their backgrounds and social characteristics. This, too, helps people open up (Krueger 1994:113).

Above all, don't lead too much and don't put words in people's mouths. In studying nutritional habits, don't ask a focus group why they eat or don't eat certain foods; do ask them to talk about what kinds of foods they like and dislike and why. In studying risky sexual behavior, don't ask, "Do you use condoms whenever you visit a prostitute?"; do ask people to talk about their experience with prostitutes and exactly what kind of sexual practices they prefer. Your job is to keep the discussion on the topic. Eventually, people will hit on the nutritional habits or the sexual acts that interest you, and you can pick up the thread from there.

Analyzing Data from Focus Groups

You can analyze focus group data with the same techniques you would use on any corpus of text: field notes, life histories, open-ended interviews, and so on. As with all large chunks of text, you have two choices for very different kinds of analysis. You can do formal content analysis, or you can do qualitative analysis. See chapter 17 (on text analysis) for more about this.

As with in-depth interviews, it's best to record (or videotape) focus groups. This is a bit tricky, though, because any audio of a focus group, whether digital or tape, is hard to understand and transcribe if two or more people talk at once. A good moderator keeps people talking one at a time. Don't hide the recorder or the microphones. Someone is sure to ask if they're being recorded, and when you tell them, "Yes"—which you must do—they're sure to wonder why they had to ask.

If you are just trying to confirm some ideas or to get a general notion of the how people feel about a topic, you can simply take notes from the tapes and work with your notes. Most focus groups, however, are transcribed. The real

power of focus groups is that they produce ethnographically rich data. Only transcription captures a significant part of that richness. But be prepared to work with a lot of information. Any single hour-and-a-half focus group can easily produce 50 pages or more of text.

Many focus groups have two staff members: a moderator and a person who does nothing but jot down the name each person who speaks and the first few words they say. This makes it easier for a transcriber to identify the voices on a tape. If you can't afford this, or if you feel that people would be uncomfortable with someone taking down their names, you can call on people by name, or mention their name when you respond to them. Things can get rolling in a focus group (that's what you want), and you'll have a tough time transcribing the tapes if you don't know who's talking.

Response Effects

Response effects are measurable differences in interview data that are predictable from characteristics of informants, interviewers, and environments. As early as 1929, Stuart Rice showed that the political orientation of interviewers can have a substantial effect on what they report their respondents told them. Rice was doing a study of derelicts in flop houses and he noticed that the men contacted by one interviewer consistently said that their down-and-out status was the result of alcohol; the men contacted by the other interviewer blamed social and economic conditions and lack of jobs. It turned out that the first interviewer was a prohibitionist and the second was a socialist (cited in Cannell and Kahn 1968:549).

Since Rice's pioneering work, hundreds of studies have been conducted on the impact of things like race, sex, age, and accent of both the interviewer and the informant; the source of funding for a project; the level of experience respondents have with interview situations; whether there is a cultural norm that encourages or discourages talking to strangers; whether the question being investigated is controversial or neutral (Cannell et al. 1979; Schuman and Presser 1981; Bradburn 1983; Schwarz 1999; Schaeffer and Presser 2003).

Katz (1942) found that middle-class interviewers got more politically conservative answers in general from lower-class respondents than did lower-class interviewers, and Robinson and Rhode (1946) found that interviewers who looked non-Jewish and had non-Jewish-sounding names were almost *four times more likely* to get anti-Semitic answers to questions about Jews than were interviewers who were Jewish looking and who had Jewish-sounding names.

Hyman and Cobb (1975) found that female interviewers who took their cars in for repairs themselves (as opposed to having their husbands do it) were more likely to have female respondents who report getting their own cars repaired. And Zehner (1970) found that when women in the United States were asked by women interviewers about premarital sex, they were more inhibited than if they were asked by men. Male respondents' answers were not affected by the gender of the interviewer.

By contrast, William Axinn (1991) found that women in Nepal were better than men as interviewers. In the Tamang Family Research Project, the female interviewers had significantly fewer "don't know" responses than did the male interviewers. Axinn supposes this might be because the survey dealt with marital and fertility histories.

Robert Aunger (1992, 2004:145–162) studied three groups of people in the Ituri forest of Zaire. The Lese and Budu are horticultural, while the Efe are foragers. Aunger wanted to know if they shared the same food avoidances. He and three assistants, two Lese men and one Budu man, interviewed a total of 65 people. Each of the respondents was interviewed twice and was asked the same 140 questions about a list of foods.

Aunger identified two types of errors in his data: forgetting and mistakes. If informants said in the first interview that they did not avoid a particular food but said in the second interview that they did avoid the food, Aunger counted the error as forgetfulness. If informants reported in interview two a different type of avoidance for a food than they'd reported in interview one, then Aunger counted this as a mistake.

Even with some missing data, Aunger had over 8,000 pairs of responses in his data (65 *pairs* of interviews, each with up to 140 responses), so he was able to look for the causes of discrepancies between interview one and interview two. About 67% of the forgetfulness errors and about 79% of the mistake errors were correlated with characteristics of informants (gender, ethnic group, age, and so on).

However, about a quarter of the variability in what informants answered to the same question at two different times was due to characteristics of the interviewers (ethnic group, gender, native language, etc.).

And consider this: About 12% of variability in forgetting was explained by interviewer experience. As the interviewers interviewed more and more informants, the informants were less likely to report "no avoidance" on interview one and some avoidance on interview two for a specific food. In other words, interviewers got better and better with practice at drawing out informants on their food avoidances.

Of the four interviewers, though, the two Lese and the Budu got much better, while the anthropologist made very little progress. Was this because of

Aunger's interviewing style, or because informants generally told the anthropologist different things than they told local interviewers, or because there is something special about informants in the Ituri forest? We'll know when we add variables to Aunger's study and repeat it in many cultures, including our own.

The Deference Effect

When people tell you what they think you want to know, in order not to offend you, that's called the **deference effect** or the **acquiescence effect**. Aunger may have experienced this in Zaire. In fact, it happens all the time, and researchers have been aware of the problem for a long, long time. In 1958, Lenski and Leggett embedded two contradictory questions in a face-to-face interview, half an hour apart. Respondents were asked whether they agreed or disagreed with the following two statements: (1) It's hardly fair to bring children into the world, the way things look for the future; (2) Children born today have a wonderful future to look forward to. Just 5% of Whites agreed with *both* statements compared to 20% of African Americans. Lenski and Leggett concluded that this was the deference effect in action: Blacks were four times more likely than Whites to agree to anything, even contradictory statements, because the interviewers were almost all white and of higher perceived status than the respondents (Lenski and Leggett 1960).

When the questions are about race, the deference effect also works for African Americans interviewing Whites. In 1989, Douglas Wilder, an African American, ran against Marshall Coleman, who is white, for the governorship of Virginia. Preelection polls showed that Wilder was far ahead, but in the end, he won by only a slim margin. When white voters were asked on the telephone whom they would vote for, they were more likely to claim Wilder as their choice if the interviewer was African American than if the interviewer was white. This effect accounted for as much as 11% of Wilder's support (Finkel et al. 1991). This finding has serious consequences for the future of election polls in the United States, as more and more elections involve competition between white and African American candidates.

Reese et al. (1986:563) tested the deference effect in a telephone survey of Anglo and Mexican American respondents. When asked specifically about their cultural preference, 58% of Hispanic respondents said they preferred Mexican American culture over other cultures, irrespective of whether the interviewer was Anglo or Hispanic. Just 9% of Anglo respondents said they preferred Mexican American culture when asked by Anglo interviewers, but 23% said they preferred Mexican American culture when asked by Hispanic interviewers.

Questions about gender and gender roles produce deference effects, too. When you ask people in the United States how most couples actually divide child care, men are more likely than women to say that men and women share this responsibility—if the interviewer is a man (Kane and McCaulay 1993:11). Do women have too much influence, just the right amount of influence, or too little influence in today's society? When asked *this* question by a male interviewer, men are more likely to say that women have *too much* influence; when asked the same question by a female interviewer, men are more likely to say that women have *too little* influence.

And similarly for women: When asked by a female interviewer, women are more likely to say that men have *too much* influence than when asked by a male interviewer (Kane and Macaulay 1993:14–15). Lueptow et al. (1990) found that women gave more liberal responses to female interviewers than to male interviewers on questions about gender roles. Men's attitudes about gender roles were, for the most part, unaffected by the gender of the interviewer—except that highly educated men gave the *most* liberal responses about gender roles to female interviewers.

“It appears,” said Lueptow et al., “that educated respondents of both sexes are shifting their answers toward the socially desirable positions they think are held by female interviewers” (p. 38). Attitudes about gender roles sure are adaptable.

Questions that aren't race related are not affected much by the race or the ethnicity of either the interviewer or the respondent. The Center for Applied Linguistics conducted a study of 1,472 bilingual children in the United States. The children were interviewed by Whites, Cuban Americans, Chicanos, Native Americans, or Chinese Americans. Weeks and Moore (1981) compared the scores obtained by white interviewers with those obtained by various ethnic interviewers and it turned out that the ethnicity of the interviewer didn't have a significant effect.

Whenever you have multiple interviewers, keep track of the race, ethnicity, and gender of the interviewer and test for response effects. Identifying sources of bias is better than not identifying them, even if you can't eliminate them. (For more on the deference effect and the social desirability effect, see Krysan and Couper 2003.)

The Third-Party-Present Effect

We sort of take it for granted that interviews are private conversations, conducted one on one, but in fact, many face-to-face interviews have at least one third party in the room, often the spouse or partner of the person being interviewed. Does this affect how people respond to questions? Sometimes it does,

and sometimes it doesn't, and there's a lot of research on when it might be a problem. Zipp and Toth (2002), for example, analyzed data from a household survey in Britain and found that when the spouses are interviewed together, they are much more likely to agree about many things—like who does what around the house—than when they are interviewed separately. Apparently, people listen to each other's answers and modify their own answers accordingly, which puts on a nice, unified face about their relationship.

As you'd expect, there is a social desirability effect when a third party is present. Casterline and Chidambaram (1984) examined data from 24 developing countries in the World Fertility Study and found that women in those countries are less likely to admit using contraception when a third party is present at the interview. Anthropologists face this situation a lot: trying to get people to talk about sensitive topics and assuring them of privacy, but unable to find the privacy for an interview.

On the other hand, Aquilino (1993) found that when their spouse is in the room, people report more marital conflict than when they are interviewed alone. They are also more likely to report that they and their spouse lived together before marriage if their spouse is in the room. Perhaps, as Mitchell (1965) suggested 40 years ago, people own up more to sensitive things like this when they know it will be obvious to their spouse that they are lying. Seems like a good thing to test. (For more on the third-party-present effect, see Blair [1979], Bradburn [1983], Hartmann [1994], Aquilino [1997], Pollner and Adams [1997], T. W. Smith [1997], Aquilino et al. [2000], and Boeije [2004]).

Threatening Questions

In general, if you are asking someone a nonthreatening question, slight changes in wording of the question won't make much difference in the answers you get. Peterson (1984) asked 1,324 people one of the following questions: (1) How old are you? (2) What is your age? (3) In what year were you born? or (4) Are you 18–24 years of age, 25–34, 35–49, 50–64, 65 or older? Then Peterson got the true ages for all the respondents from reliable records.

There was no significant difference in the accuracy of the answers obtained with the four questions. (However, almost 10% of respondents refused to answer question 1, while only 1% refused to answer question 4, and this difference *is* significant.)

On the other hand, if you ask people about their alcohol consumption, or whether they ever shoplifted when they were children, or whether they have family members who have had mental illness, or how many sexual partners they've had, then even small changes in the wording can have significant

effects on informants' responses. And asking about *other people's* sexual behavior, by the way, can produce dramatically different results. Katz and Naré (2002) asked 1,973 single Muslim women between the ages of 15 and 24 in Dakar, Senegal, if they had ever been pregnant. Three percent of the women said they had. But 25% of the same women said that at least one of their *three closest friends* had been pregnant—more than eight times what they reported about themselves. (See Wiederman et al. [1994], Catania et al. [1996], Gribble et al. [1999], and Hewitt [2002] for work on how to increase response to questions about sexual behavior. For more on threatening questions in general and the use of the three-closest-friends technique, see Bradburn [1983:147–151]; on improving response to threatening questions, see Bradburn et al. 1978 and Bradburn, Sudman et al. 1979. See Johnston and Walton [1995] on the use of computer-assisted self-interviewing for asking sensitive questions. And see below for more on computer-assisted interviewing.)

The Expectancy Effect

In 1966, Robert Rosenthal conducted an experiment. At the beginning of the school year, he told some teachers at a school that the children they were about to get had tested out as “spurters.” That is, according to tests, he said, those particular children were expected to make significant gains in their academic scores during the coming year. Sure enough, those children did improve dramatically—which was really interesting, because Rosenthal had matched the “spurter” children and teachers at random.

The results, published in a widely read book called *Pygmalion in the Classroom* (Rosenthal and Jacobson 1968) established once and for all what experimental researchers across the behavioral sciences had long suspected. There is an **expectancy effect**. The expectancy effect is “the tendency for experimenters to obtain results they expect, not simply because they have correctly anticipated nature's response but rather because they have helped to shape that response through their expectations” (Rosenthal and Rubin 1978:377).

In 1978, Rosenthal and Rubin reported on the “first 345 studies” that were generated by the discovery of the expectancy effect, and research continues on this problem (see Rosenthal 2002). The effect is largest in animal studies (perhaps because there is no danger that animals will go into print rejecting findings from experiments on them), but it is likely in all experiments on people. As Rosenthal's first study proved, the effect extends to teachers, managers, therapists—anyone who makes a living creating changes in the behavior of others.

Expectancy is different from distortion. The distortion effect comes from seeing what you want to see, even when it's not there. The expectancy effect

involves creating the objective results we want to see. We don't distort results to conform to our expectations as much as we make the expectations come true.

Strictly speaking, then, the expectancy effect is not a response effect at all. But for fieldworkers, it is an important effect to keep in mind. If you are studying a small community, or a neighborhood in a city, or a hospital or clinic for a year or more, interacting daily with a few key informants, your own behavior can affect theirs in subtle (and not so subtle) ways, and vice versa. Don't be surprised if you find your own behavior changing over time in relation to key informants.

Accuracy

Even when people tell you what they think is the absolute truth, there is still the question of whether the information they give you is accurate.

A lot of research—ethnographic and survey research alike—is about mapping opinions and attitudes. When people tell you that they *approve of* how the chief is handling negotiations for their village's resettlement, or when they tell you that they *prefer* a particular brand of beer to some other brand, they're talking about internal states. You pretty much have to take their word for such things.

But when we ask people to tell us about their actual behavior (How many times did you take your baby to the clinic last month? How many times last year did you visit your mother's village?), or about their environmental circumstances (How many hectares of land do you have in maize? How many meters is it from your house to the well?), we can't just assume informant accuracy.

We see reports of behavior in our local newspapers all the time: College students today are binge drinking more than they did 5 years ago. Americans are going to church more often than they did a decade ago. In back of *findings* like these are *questions* like these:

Circle one answer:

How many times last month did you consume five or more beers or other alcoholic drinks in a single day?

Never

Once

Twice

Three times

More than three times

How often do you go to church?

Never

Very occasionally

About once a month

About once a week

More than once a week

La Pierre Discovers the Problem

We've known for a long time that we should be suspicious of this kind of data. From 1930 to 1932, Richard La Pierre, accompanied by a Chinese couple, crisscrossed the United States, twice, by car. The threesome covered about 10,000 miles, stopping at 184 restaurants and 66 hotels. And they kept records. There was a lot of prejudice against Chinese in those days, but they were not refused service in a single restaurant and just one hotel turned them away (La Pierre 1934).

Six months after the experiment ended, La Pierre sent a questionnaire to each of the 250 establishments where the group had stopped. One of the things he asked was: "Will you accept members of the Chinese race as guests?" Ninety-two percent—230 out of 250—replied "No."

By today's standards, La Pierre's experiment was crude. He could have surveyed a control group—a second set of 250 establishments that they hadn't patronized but that were in the same towns where they'd stopped. With self-administered questionnaires, he couldn't be sure that the people who answered the survey (and who claimed that they wouldn't serve Chinese) were the same ones who had actually served the threesome. And La Pierre didn't mention in his survey that the Chinese couple would be accompanied by a white man.

Still, La Pierre's experiment was terrific for its time. It made clear that what people say they do (or would do) is not a proxy for what they actually do or will do (see Deutscher 1973). This basic finding shows up in the most unlikely (we would have thought) places: In the 1961 census of Addis Ababa, Ethiopia, 23% of the women underreported the *number of their children!* Apparently, people there didn't count babies who die before reaching the age of two (Pausewang 1973:65).

Why People Are Inaccurate Reporters of Their Own Behavior

People are inaccurate reporters of their own behavior for many reasons. Here are four:

1. Once people agree to be interviewed, they have a personal stake in the process and usually try to answer all your questions—whether they understand what you're after or not.

2. Human memory is fragile, although it's clearly easier to remember some things than others.

Cannell et al. (1961) found that the ability to remember a stay in the hospital is related to the length of the stay, the severity of the illness that lands you there, and whether or not surgery is involved. It's also strongly related to the length of time since discharge. Cannell and Fowler (1965) found that people report accurately 90% of all overnight hospital stays that happened 6 months or less before being interviewed.

It's easy for people to remember a rare event, like surgery, that occurred recently. But, as Sudman and Schwarz (1989) point out, if you ask people to think about some common behavior going back months at a time, they probably use estimation rules. When Sudman and Schwartz asked people "How many [sticks] [cans] of deodorant did you buy in the last six months?" they started thinking: "Well, I usually buy deodorant about twice a month in the summer, and about once a month the rest of the year. It's now October, so I suppose I must have bought 10 deodorants over the last six months." And then they say, "10," and that's what you write down.

3. Interviews are social encounters. People manipulate those encounters to whatever they think is their advantage.

Adolescent boys tend to exaggerate, and adolescent girls tend to minimize, reports of their own sexual experience (see Catania et al. 1996). Expect people to overreport socially desirable behavior and to underreport socially undesirable behavior. (See deMaio [1984] for a review of the social desirability effect.)

4. People can't count a lot of behaviors, so they use rules of inference.

In some situations, they invoke D'Andrade's "what goes with what" rule (1974) and report what they *suppose* must have happened, rather than what they actually saw. Freeman et al. (1987) asked people in their department to report on who attended a particular colloquium. People who were *usually* at the department colloquium were mentioned as having attended the particular colloquium—even by those who hadn't attended (and see Shweder and D'Andrade 1980).

Reducing Errors: Jogging Informants' Memories

Sudman and Bradburn (1974) distinguish two types of memory errors: simply forgetting and reporting that something happened a month ago when it

really happened two months ago. The latter error is called **forward telescoping** (backward telescoping is rare).

Here are four things you can do to increase the accuracy of self-reported behavior.

1. **Cued recall.** In cued recall, people either consult records to jog their memories or you ask them questions that cue them about specific behaviors. For example, if you're collecting life histories, college transcripts will help people remember events and people from their time at school. Credit card statements and long-distance phone bills help people retrace their steps and remember events, places, and people they met along the way. Still . . . Horn (1960) asked people to report their bank balance. Of those who did not consult their bankbooks, just 31% reported correctly. But those who consulted their records didn't do that much better. Only 47% reported correctly (reported in Bradburn 1983:309).

Event calendars are particularly useful in societies where there are no written records. Leslie et al. (1999:375–378), for example, developed an event calendar for the Ngisonyoka section of the South Turkana pastoralists in northwestern Kenya. The Turkana name their seasons rather than their years. Based on many interviews between 1983 and 1984, Leslie et al. were able to build up a list of 143 major events associated with seasons between 1905 and 1992. Events include things like “no hump” in 1961 (it was so dry that the camels' humps shrank), “bulls” in 1942 (when their bulls were taken to pay a poll tax), and “rescue” in 1978 (when rains came). This painstaking work has made it possible for many researchers to gather demographic and other life history data from the Ngisonyoka Turkana. (For more on event calendars in life histories, see Freedman et al. 1988, Kessler and Wethington 1991, Caspi et al. 1996, and Belli 1998.)

Brewer and Garrett (2001) found that five kinds of questions can dramatically increase the recall of sex partners and drug injection partners. They gave people alphabetic cues, location cues, network cues, role cues, and timeline cues. After asking people to list their sex partners and/or drug injection partners, they asked them the following questions:

1. Alphabetic cues. “I am going to go through the letters of the alphabet one at a time. As I say each letter, think of all the people you know whose name begins with that letter. The names could be first names, nicknames, or last names. Let me know if any of these are people you had sex/injected drugs with in the last year but might not have mentioned earlier.”
2. Location cues. “I have here a list of different kinds of locations or places where people have sex/inject drugs with other people or meet people who they later have sex/inject drugs with. As I say each location, think of all of the times you have had sex/injected drugs there, or met people there in the last year. Focus on all the people you interacted with at these locations. Let me know if any of these are people you had sex/injected drugs with but might not have mentioned earlier.”
3. Network cues. “I am going to read back again the list of people you mentioned

earlier. This time, as I say each person, think of all the other people who know, hang out, or interact with that person. Let me know if any of these are people you had sex/injected drugs with in the last year but might not have mentioned earlier.”

4. **Role cues.** “I have here a list of different kinds of relationships people have with the persons they have sex/inject drugs with. As I say each type of role relationship, think of all of the people you know that you have that kind of relationship with. Let me know if any of these are people you had sex/injected drugs with in the last year but might not have mentioned earlier.”
5. **Timeline cues.** “We’re going to map out where you’ve been and what you’ve been doing the last year. Then we will go through this timeline and see if there are other people you have had sex/injected drugs with during this period. As we are making this timeline, if any additional people you have had sex/injected drugs with during this period come to mind, please tell me.”

Asking these five questions together increased the number of sex partners recalled by 40% and the number of drug injection partners by 123% (Brewer and Garrett 2001:672; the questions are from <http://faculty.washington.edu/ddbrewer/trevinstr.htm>).

2. **Aided recall.** In this technique, you hand people a list of possible answers to a question and ask them to choose among them. Aided recall increases the number of events recalled, but also appears to increase the telescoping effect (Bradburn 1983:309). Aided recall is particularly effective in interviewing the elderly (Jobe et al. 1996).

In studies where you interview people more than once, another form of aided recall is to remind people what they said last time in answer to a question and then ask them about their behavior since their last report. This corrects for telescoping but does not increase the number of events recalled.

3. **Landmarks.** Here, you try to establish a personal milestone—like killing your first peccary, going through clitoridectomy, burying your mother, becoming a grandparent—and asking people to report on what has happened since then.

Loftus and Marburger (1983) found that landmarks help reduce forward telescoping. The title of their articles says it all: “Since the Eruption of Mt. St. Helens, Has Anyone Beaten You Up? Improving the Accuracy of Retrospective Reports with Landmark Events.” Means et al. (1989) asked people to recall landmark events in their lives going back 18 months from the time of the interview. Once the list of personal landmark events was established, people were better able to recall hospitalizations and other health-related events.

4. **Restricted time.** Sudman and Schwarz (1989) advocate keeping the recall period short in order to increase recall accuracy. They asked people: “How many times have you been out to a restaurant in the last three months?” and “How many times have you been out to a restaurant in the last month?” The per-month average for the 1-month question was 55% greater than the per-month average for

the 3-month question. The assumption here is that increasing the amount of the behavior reported also increases its accuracy.

The Social Desirability Effect

Hadaway et al. (1998) went to a large Protestant church and found 115 people in attendance at the Sunday school. On Monday morning, when Hadaway et al. polled the whole church membership, 181 people claimed to have been in Sunday school the previous day. Head-count experiments like this one typically produce estimates of church attendance that are 55%–59% of what people report (T. W. Smith 1998).

This **social desirability** effect is influenced by the way you ask the question. Major surveys, like the Gallup Poll, ask something like: “How often do you attend religious services?” Then they give the people choices like “once a week, once a month, seldom, never.” Presser and Stinson (1998) asked people on Monday to list everything they had done from “midnight Saturday to midnight last night.” When they asked the question this way, 29% of respondents said that they had gone to church. Asking “How often do you go to church?” produced estimates of 37%–45%. (This is a 28%–55% *difference* in reported behavior and is statistically very significant.)

Informant accuracy remains a major problem. Gary Wells and his colleagues (2003) showed a video of a staged crime to 253 students. Then they showed the students a photo lineup of six people and asked the students to pick out the culprit. Every single student picked one of the six photos, but there was a small problem: the culprit wasn’t in the six photos. We need a lot more research about the rules of inference that people use when they respond to questions about where they’ve been, who they were with, and what they were doing.

10



Structured Interviewing I: Questionnaires

This is the first of two chapters about **structured interviews**. In a structured interview, each informant or respondent is exposed to the same stimuli. The stimuli are often questions, but they may also be carefully constructed vignettes, lists of words or photos, clips of music or video, a table full of physical artifacts, or a garden full of plants. The idea in structured interviewing is always the same: to control the input that triggers people's responses so that their output can be reliably compared.

I'll cover two broad categories of methods for structured interviewing: **questionnaires** and a range of methods used in **cultural domain analysis**. We begin in this chapter with questionnaires and survey research. I review some of the important lessons concerning the wording of questions, the format of questionnaires, the management of survey projects, and the maximizing of response rates. (Refer to chapter 9 again for more discussion of response effects.)

At the end of this chapter, I'll introduce you to some of the interesting and unusual methods that people are using to get at complex and/or touchy topics. In chapter 11, I'll introduce you to the structured data-gathering methods for cultural domain analysis.

Questionnaires and Survey Research

Survey research goes back over 200 years (take a look at John Howard's monumental 1973 [1792] survey of British prisons), but it really took off in

the mid-1930s when quota sampling was first applied to voting behavior studies and to helping advertisers target consumer messages. Over the years, government agencies in all the industrialized countries have developed an insatiable appetite for information about various “target populations” (poor people, users of public housing, users of private health care, etc.). Japan developed an indigenous survey research industry soon after World War II, and India, South Korea, Jamaica, Greece, Mexico, and many other countries have since developed their own survey research capabilities.

Anthropologists are finding more and more that good survey technique can add a lot of value to ethnography. In the 1970s, Sylvia Scribner and Michael Cole studied literacy among the Vai of Liberia. Some Vai are literate in English, others are literate in Arabic, and some adult Vai men use an indigenous script for writing letters. As part of their project, Scribner and Cole ran a survey with 650 respondents. Michael Smith, the cultural anthropologist on their team, was skeptical about using this method with the Vai. He wrote the project leaders about his experience in administering the survey there:

I was surprised when I first saw how long it [the questionnaire] was. I didn't think that anyone would sit down for long enough to answer it, or, if they did, that they would answer it seriously. . . . Well, I was wrong—and it fascinates me why the Vai should, in the busiest season of the year—during two of the worst farming years one could have picked . . . spend a lot of time answering questions which had little to do with the essential business at hand. . . . Not only did the majority of people eventually come, but when they got there they answered with great deliberation. How many times does one remember someone saying, “I don't know, but I'll come back and tell you when I've checked with so-and-so.” (Scribner and Cole 1981:47)

A lot has changed in the last 30 years. Today, most anthropologists use questionnaires as one of their research tools.

The Computer Revolution in Survey Research

There are three methods for collecting survey questionnaire data: (1) personal, **face-to-face** interviews, (2) **self-administered** questionnaires, and (3) **telephone** interviews. All three of these methods can be either assisted by, or fully automated with, computers.

The computer revolution in survey research began in the 1970s with the development of software for CATI, or “computer-assisted telephone interviewing.” By 1980, CATI software had transformed the telephone survey industry (Fink 1983). With CATI software, you program a set of survey ques-

tions and then let the computer do the dialing. Interviewers sit at their computers, wearing telephone headsets, and when a respondent agrees to be interviewed, they read the questions from the screen. With the kind of fixed-choice questions that are typical in surveys, interviewers only have to click a box on the screen to put in the respondent's answer to each question. For open-ended questions, respondents talk and the interviewer types in the response.

CASI stands for "computer-assisted self-administered interview." People sit at a computer and answer questions on their own, just like they would if they received a questionnaire in the mail. People can come to a central place to take a CASI survey or you can send them a disk in the mail that they can plug into their own computer . . . or you can even set up the survey on the web and people can take it from any Internet connection. (For more on disk-by-mail surveys, see Van Hattum and de Leeuw [1999]. For more on CASI, see de Leeuw and Nicholls [1996], Nicholls et al. [1997], and de Leeuw et al. [2003].) People take very quickly to computer-based interviews and often find them to be a lot of fun. Fun is good because it cuts down on fatigue. Fatigue is bad because it sends respondents into robot mode and they stop thinking about their answers (O'Brien and Dugdale 1978; Barnes et al. 1995). I ran a computer-based interview in 1988 in a study comparing the social networks of people in Mexico City and Jacksonville, Florida. One member of our team, Christopher McCarty, programmed a laptop to ask respondents in both cities about their acquaintanceship networks. Few people in Jacksonville and almost no one in Mexico City had ever seen a computer, much less one of those clunky lugables that passed for laptops then. But our respondents said they enjoyed the experience. "Wow, this is like some kind of computer game," one respondent said.

Today, of course, the technology is wildly better and fieldworkers are running computer-assisted interview surveys all over the world. Hewett et al. (2004) used A-CASI technology—for "audio, computer-assisted, self-administered interview"—in a study of 1,293 adolescents in rural and urban Kenya about very sensitive issues, like sexual behavior, drug and alcohol use, and abortion. With A-CASI, the respondent listens to the questions through headphones and types in his or her answers. The computer—a digitized voice—asks the questions, waits for the answers, and moves on. In the Kenya study, Hewett et al. used yes/no and multiple choice questions and had people punch in their responses on an external keypad. The research team had to replace a few keypads and they had some cases of battery failure, but overall, they report that the computers worked well, that only 2% of the respondents had trouble with the equipment (even though most of them had never seen a computer) and that people liked the format (*ibid.*:322–324).

This doesn't mean that computers are going to replace live interviewers any

time soon. Computers-as-interviewers are fine when the questions are very clear and people don't need a lot of extra information. Suppose you ask: "Did you go to the doctor last week?" and the informant responds: "What do you mean by doctor?" She may have gone to a free-standing clinic and seen a nurse practitioner or a physician's assistant. She probably wants to know if this counts as "going to the doctor." (For more on audio CASI, see Aquilino et al. [2000], Newman et al. [2002], and Couper et al. [2003].)

CAPI software supports "computer-assisted personal interviewing." With this technology, you build your interview on a laptop or a handheld computer. The computer prompts you with each question, suggests probes, and lets you enter the data as you go. CAPI doesn't replace you as the interviewer, it just makes it easier for you to enter and manage the data. Easier is better, and not just because it saves time. It also reduces errors in the data. When you write down data by hand in the field, you are bound to make some errors. When you input those data into a computer, you're bound to make some more errors. The fewer times you have to handle and transfer data, the better.

There are now MCAPI (mobile CAPI) programs. This technology is particularly good for anthropologists. You program a survey into a handheld computer. As you interview people in the field, you punch in the data on the fly. Clarence Gravlee (2002a) used this method to collect data on lifestyle and blood pressure from 100 people in Puerto Rico. His interviews had 268 multiple choice, yes/no, and open-ended questions, and took over an hour to conduct, but when he got home each night from his fieldwork, he had the day's data in the computer. He had to go over the data carefully to catch any errors, but every time you record data there is a risk of making mistakes. But cutting down on the number of times you have to handle data means fewer mistakes. (See Greene [2001] on using handheld computers in fieldwork and see appendix F for websites on handhelds.)

Internet-Based Surveys

The latest advance in computer-assisted survey research is fully automated interviews on the Internet. Many companies sell software for building Internet surveys, and there are hundreds of examples out there of social research based on these surveys. (For a sampling of studies, go to <http://psych.hanover.edu/research/exponnet.html>.)

As with any format, there are pluses and minuses to Internet surveys. On the plus side, the Internet makes it easy to recruit respondents in otherwise hard-to-reach groups. To study gay Latino men, Ross et al. (2004) placed about 47 million banner ads on gay-themed websites inviting potential respon-

dents for a university-sponsored study. The ads produced about 33,000 clicks, 1,742 men who started the survey, and 1,026 men who finished it.

The response rate to Internet surveys and mailed surveys can be comparable, if you send potential respondents a note by regular mail telling them about the study (Kaplowitz et al. 2004). But this assumes that you have a sampling frame with the names and addresses of the people you want to interview. In 2000, my colleagues and I tried to do a national survey of people who met two simultaneous criteria: They had access to the Internet and they had purchased a new car in the last 2 years or were in the market for a car now. There's no sampling frame of such people, so we decided to do a national RDD (random-digit-dialing) survey to find people who met our criteria. Then, we'd offer them \$25 to participate in an Internet survey, and, if they agreed, we'd give them the URL of the survey and a PIN.

We made 11,006 calls and contacted 2,176 people. That's about right for RDD surveys. The rest of the numbers either didn't answer, or were businesses, or there was only a child at home, etc. Of the 2,176 people we contacted, 910 (45%) were eligible for the web survey. Of them, 136 went to the survey site and entered their PIN, and of them, 68 completed the survey. The data from those 68 people were excellent, but it took an awful lot of work for a purposive (nonrepresentative) sample of 68 people.

In fact, Internet surveys are used mostly in studies that don't require representative samples. In 2003, only about 80% of households in the United States with college graduates had access to the Internet and that dropped to 24% in households with less than a high school education (SAUS 2004–2005, table 1152). (About 43 million adults had less than a high school education in the United States in 2003.) By contrast, over 95% of households in the United States have a telephone. Even in households with less than \$5,000 annual income, the penetration of telephones is around 82% (see the section on telephone surveys, below).

Still, Internet surveys hold great promise for anthropologists. There are Internet points in the most out-of-the-way places today, which means that we can continue to interview our informants between trips to the field.

Advantages and Disadvantages of Survey Formats

Each major data-collection method—face-to-face, self-administered, and telephone interview—has its advantages and disadvantages. There is no conclusive evidence that one method of administering questionnaires is better, overall, than the others. Your choice of a method will depend on your own

calculus of things like cost, convenience, and the nature of the questions you are asking.

Personal, Face-to-Face Interviews

Face-to-face administration of questionnaires has major advantages, but it also has some disadvantages as well.

ADVANTAGES OF FACE-TO FACE INTERVIEWS

1. They can be used with people who could not otherwise provide information—respondents who are illiterate or nonliterate, blind, bedridden, or very old, for example.
2. If a respondent doesn't understand a question in a personal interview, you can fill in, and, if you sense that the respondent is not answering fully, you can probe for more complete data.

Conventional wisdom in survey research is that each respondent has to hear exactly the same question. In practice, this means not engaging in conversation with people who ask for more information about a particular item on a survey. Not responding to requests for more information might mean sacrificing validity for reliability. There is now evidence that a more conversational style produces more accurate data, especially when respondents really need to get clarifications on unclear concepts (Schober and Conrad 1997; Krosnick 1999).

So, carry a notebook that tells you exactly how to respond when people ask you to clarify an unfamiliar term. If you use more than one interviewer, be sure each of them carries a copy of the same notebook. Good interview schedules are pretested to eliminate terms that are unfamiliar to intended respondents. Still, there is always someone who asks: "What do you mean by 'income'?" or "How much is 'a lot'?"

3. You can use several different data collection techniques with the same respondent in a face-to-face survey interview. Part of the interview can consist of open-ended questions; another part may require the use of visual aids, such as graphs or cue cards; and in still another, you might hand the respondent a self-administered questionnaire booklet and stand by to help clarify potentially ambiguous items. This is a useful technique for asking really sensitive questions in a face-to-face interview.
4. Personal interviews at home can be much longer than telephone or self-administered questionnaires. An hour-long personal interview is relatively easy, and even 2- and 3-hour interviews are common. It is next to impossible to get respon-

dents to devote 2 hours to filling out a questionnaire that shows up in the mail, unless you are prepared to pay well for their time; and it requires exceptional skill to keep a telephone interview going for more than 20 minutes, unless respondents are personally interested in the topic (Holbrook et al. 2003). Note, though, that **street-intercept** or **mall-intercept** interviews (where you interview people on the fly), while face to face, usually have to be very quick.

5. Face-to-face respondents get one question at a time and can't flip through the questionnaire to see what's coming. If you design an interview to start with general questions (how people feel about using new technologies at work, for example) and move on to specific questions (how people feel about using a particular new technology), then you really don't want people flipping ahead.
6. With face-to-face interviews, you know who answers the questions.

DISADVANTAGES OF FACE-TO-FACE INTERVIEWS

1. They are intrusive and **reactive**. It takes a lot of skill to administer a questionnaire without subtly telling the respondent how you hope he or she will answer your questions. Other methods of administration of questionnaires may be impersonal, but that's not necessarily bad, especially if you've done the ethnography and have developed a set of fixed-choice questions for a questionnaire. Furthermore, the problem of reactivity increases when more than one interviewer is involved in a project. Making it easy for interviewers to deliver the same questions to all respondents is a plus.

2. Personal interviews are costly in both time and money. In addition to the time spent in interviewing people, locating respondents in a representative sample may require going back several times. In urban research especially, count on making up to half a dozen callbacks to get the really hard-to-find respondents.

It's important to make all those callbacks in order to land the hard-to-get interviews. Survey researchers sometimes use the **sampling by convenient replacement** technique—going next door or down the block and picking up a replacement for an interviewee who happens not to be home when you show up. As I mentioned in chapter 6, this tends to homogenize your sample and make it less and less representative of all the variation in the population you're studying.

3. If you are working alone, without assistants, in an area that lacks good roads, don't plan on doing more than around 200 survey interviews in a year. If you're working in major cities in Europe or North America you can do more, but it gets really, really tough to maintain a consistent, positive attitude long before you get to the 200th interview. With mailed and telephone questionnaires, you can survey thousands of respondents.
4. Personal interview surveys conducted by lone researchers over a long period of time run the risk of being overtaken by events. A war breaks out, a volcano erupts, or the government decides to cancel elections and imprison the opposition. It sounds dramatic, but these sorts of things are actually quite common across the world. Far less dramatic events can make the responses of the last 100

people you interview radically different from those of the first 100 to the same questions. If you conduct a questionnaire survey over a long period of time in the field, it is a good idea to reinterview your first few respondents and check the stability (reliability) of their reports.

Interviewer-Absent Self-Administered Questionnaires

Mailed questionnaires, questionnaires dropped off at people's homes or where they work, questionnaires that people pick up and take home with them, and questionnaires that people take on the Internet—all these are interviewer-absent, self-administered instruments for collecting survey data.

These truly **self-administered questionnaires** have some clear advantages and disadvantages.

ADVANTAGES OF SELF-ADMINISTERED QUESTIONNAIRES

1. Mailed questionnaires (whether paper or disk) puts the post office to work for you in finding respondents. If you cannot use the mail (because sampling frames are unavailable, or because you cannot expect people to respond, or because you are in a country where mail service is unreliable), you can use cluster and area sampling (see chapter 6), combined with the **drop-and-collect** technique. This involves leaving a questionnaire with a respondent and going back later to pick it up. Ibeh and Brock (2004) used this in their study of company managers in Nigeria. The standard response rate for questionnaires mailed to busy executives in Sub-Saharan Africa is around 36%. Using the drop-and-collect technique, Ibeh and Brock achieved a nearly 60% response rate. With both mailed surveys and the drop-and-collect method, self-administered questionnaires allow a single researcher to gather data from a large, representative sample of respondents, at relatively low cost per datum.
2. All respondents get the same questions with a self-administered questionnaire. There is no worry about interviewer bias.
3. You can ask a *bit* more complex questions with a self-administered paper questionnaire than you can in a personal interview. Questions that involve a long list of response categories, or that require a lot of background data are hard to follow orally, but are often interesting to respondents if worded right.

But for really complex questions, you're better off with CASI. In computer-assisted self-administered interviews, people don't have to think about any convoluted instructions at all—instructions like: "Have you ever had hepatitis? If not, then skip to question 42." Later, after the respondent finishes a series of questions about her bout with hepatitis, the questionnaire says: "Now

return to question 40.” With a CASI, the computer does all the work and the respondent can focus on responding.

4. You can ask long **batteries** of otherwise boring questions on self-administered questionnaires that you just couldn’t get away with in a personal interview. Look at figure 10.1. Imagine trying to ask someone to sit still while you recited, say, 30 items and asked for their response. And again, computer-assisted interviewing is even better at this.

Here is a list of things that people say they’d like to see in their high school. For each item, check how you feel this high school is doing.

	WELL	OK	POORLY	DON'T KNOW
1. High-quality instruction	_____	_____	_____	_____
2. Good pay for teachers	_____	_____	_____	_____
3. Good mix of sports and academics	_____	_____	_____	_____
4. Preparation for college entrance exams	_____	_____	_____	_____
5. Safety	_____	_____	_____	_____
6. Music program	_____	_____	_____	_____
7. Good textbooks	_____	_____	_____	_____
⋮				
⋮				

Figure 10.1. A battery item in a questionnaire. Batteries can consist of many items.

5. There are question-order effects and acquiescence effects in self-administered interviews, just as there are in other instruments. (Acquiescence is the tendency for some people to respond to anything, even if they don’t know the answer, just to satisfy the questioner.) But **response effects**, based on features of the interviewer, are not a problem. Questions about sexual behavior (including family planning) and about attitudes toward women or men or members of particular ethnic/racial groups are particularly susceptible to this problem. The perceived sexual orientation of the interviewer, for example, affects how supportive respondents are of homosexuality (Kemph and Kasser 1996).
6. Some people are more willing to report socially undesirable behaviors and traits in self-administered questionnaires (and in telephone interviews) than they are in face-to-face interviews (Aquilino 1994; de Leeuw et al. 1995; Tourangeau and Smith 1996). Peterson et al. (1996) randomly assigned two groups of 57 Swedish Army veterans to fill out the Beck’s Depression Inventory (Beck et al. 1961).

One group used the pencil-and-paper version, while the other used a computer-based version. Those who used the computer-based version had significantly higher mean scores on really sensitive questions about depression.

In self-administered interviews, people aren't trying to impress anyone, and anonymity provides a sense of security, which produces more reports of things like premarital sexual experiences, constipation, arrest records, alcohol dependency, interpersonal violence, and so on (Hochstim 1967; Bradburn 1983).

This does *not* mean that *more* reporting of behavior means more *accurate* reporting. We know better than that now. But, as I've said before, more is usually better than less. If Chicanos report spending 12 hours per week in conversation with their families at home, while Anglos (as white, non-Hispanic Americans are known in the American Southwest) report spending 4 hours, I wouldn't want to bet that Chicanos *really* spend 12 hours, on average, or that Anglos *really* spend 4 hours, on average, talking to their families. But I'd find the fact that Chicanos reported spending three times as much time talking with their families pretty interesting.

DISADVANTAGES OF SELF-ADMINISTERED QUESTIONNAIRES

Despite these advantages, there are some hefty disadvantages to self-administered questionnaires.

1. You have no control over how people interpret questions on a self-administered instrument, whether the questionnaire is delivered on paper or on a computer or over the Internet. There is always the danger that, no matter how much background work you do, no matter how hard you try to produce culturally correct questions, respondents will be forced into making culturally inappropriate choices in closed-ended questionnaires. If the questionnaire is self-administered, you can't answer people's questions about what a particular item means.
2. If you are not working in a highly industrialized nation, or if you are not prepared to use Dillman's Total Design Method (discussed below), you are likely to see response rates of 20%–30% from mailed questionnaires. It is entirely reasonable to analyze the data statistically and to offer conclusions about the correlations among variables for those who responded to your survey. But response rates like these are unacceptable for drawing conclusions about larger populations. CASI and audio CASI studies are based on real visits with people, in the field. Response rates for those forms of self-administered questionnaires can be very high. In that study that Hewett et al. did in Kenya, they had a response rate of over 80% (2004:328).
3. Even if a mailed questionnaire is returned, you can't be sure that the respondent who received it is the person who filled it out. And similarly for Internet and e-mail questionnaires.

4. Mailed questionnaires are prone to serious sampling problems. Sampling frames of addresses are almost always flawed, sometimes very badly. If you use a phone book to select a sample, you miss all those people who don't have phones or who choose not to list their numbers. Face-to-face administration of questionnaires is often based on an area cluster sample, with random selection of households within each cluster. This is a much more powerful sampling design than most mailed questionnaire surveys can muster.
5. In some cases, you may want respondents to answer a question without their knowing what's coming next. This is impossible in a self-administered paper questionnaire, but it's not a problem in CASI and audio CASI studies.
6. Self-administered paper and CASI questionnaires are simply not useful for studying nonliterate or illiterate populations, or people who can't use a keyboard. This problem will eventually be solved by voice recognition software, but we're just at the beginning of that particular revolution.

Telephone Interviews

Once upon a time, telephone surveys were considered a poor substitute for face-to-face surveys. Today, telephone interviewing is the most widely used method of gathering survey data across the industrialized nations of the world where so many households have their own phones. Administering questionnaires by phone has some very important advantages.

ADVANTAGES OF TELEPHONE INTERVIEWS

1. Research has shown that, in the United States at least, answers to many different kinds of questions asked over the phone are as valid as those to questions asked in person or through the mail (Dillman 1978).
2. Phone interviews have the impersonal quality of self-administered questionnaires and the personal quality of face-to-face interviews. So, telephone surveys are unintimidating (like self-administered questionnaires), but allow interviewers to probe or to answer questions dealing with ambiguity of items (just like they can in personal interviews).
3. Telephone interviewing is inexpensive and convenient to do. It's not without effort, though. Professional survey organizations routinely do at least three callbacks to numbers that don't answer, and many survey researchers insist on 10 callbacks to make sure that they get an unbiased sample. As it is, in most telephone surveys, you can expect 30%–40% refusals. You can also expect nearly 100% sample completion, because it's relatively easy to replace refusers with people who will cooperate. But remember to keep track of the refusal rate and to make an extra effort to get at least some of the refusers to respond so you can test whether cooperators are a biased sample.
4. Using **random digit dialing** (RDD), you can reach almost everyone who has a phone. In the United States, that means you can reach almost everybody. One

survey found that 28% of completed interviews using RDD were with people who had unlisted phone numbers (Taylor 1997:424). There are huge regional differences, though, in the availability of telephones (see below).

5. Unless you do all your own interviewing, interviewer bias is an ever-present problem in survey research. It is relatively easy to monitor the quality of telephone interviewers' work by having them come to a central place to conduct their operation. (But if you don't monitor the performance of telephone interviewers, you invite cheating. See below, in the section on the disadvantages of telephone interviewing.)
6. There is no reaction to the *appearance* of the interviewer in telephone surveys, although respondents *do* react to accents and speech patterns of interviewers. Oskenberg et al. (1986) found that telephone interviewers who had the lowest refusal rates had higher-pitched, louder, and clearer voices. And, as with all types of interviews, there are gender-of-interviewer and race-of-interviewer effects in telephone interviews, too. Respondents try to figure out the race or ethnicity of the interviewer and then tailor responses accordingly.

In the National Black Election Study, 872 African Americans were polled before and after the 1984 presidential election. Since interviewers were assigned randomly to respondents, some people were interviewed by a white person before the election and an African American after the election. And vice versa: Some people were interviewed by an African American before the election and a white person on the second wave.

Darren Davis (1997) looked at data from this natural experiment. When African American interviewers in the preelection polls were replaced by white interviewers in the postelection surveys, African Americans were more likely to say that Blacks don't have the power to change things, that Blacks can't make a difference in local or national elections, that Blacks cannot form their own political party, and that Whites are not responsible for keeping Blacks down—very powerful evidence of a race-of-interviewer effect.

7. Telephone interviewing is safe. You can talk on the phone to people who live in urban neighborhoods where many professional interviewers (most of whom are women) would prefer not to go. Telephones also get you past doormen and other people who run interference for the rich.

DISADVANTAGES OF TELEPHONE INTERVIEWING

The disadvantages of telephone surveys are obvious.

1. If you are doing research in Haiti or Bolivia or elsewhere in the developing world, telephone surveys are out of the question, except for some urban centers, and then only if your research is about relatively well-off people.

Even in highly industrialized nations, not everyone has a telephone. About 95% of all households in the United States have telephones. This makes *national* surveys a cinch to do and highly reliable. But the distribution of telephones is uneven, which makes some *local* surveys impossible to do by phone. Almost every household in the United States with a median annual income of at least \$60,000 has a phone, but only 82% of households with median annual incomes below \$5,000 have a phone.

2. Telephone interviews must be relatively short, or people will hang up. There is some evidence that once people agree to give you their time in a telephone interview, you can keep them on the line for a remarkably long time (up to an hour) by developing special “phone personality” traits. Generally, however, you should not plan a telephone interview that lasts for more than 20 minutes.
3. Random-digit-dialing phone surveys are big business, and many people are turned off by them. It is becoming increasingly difficult for telephone survey research organizations to complete interviews. It may take thousands of phone calls to get a few hundred interviews. No one knows yet how this may be compromising the validity of the data collected in RDD surveys.
4. And finally, this: It has long been known that, in an unknown percentage of occasions, hired interviewers willfully produce inaccurate data. When an interviewer who is paid by the completed interview finds a respondent not at home, the temptation is to fill in the interview and get on to the next respondent. This saves a lot of calling back, and introduces garbage into the data.

Unless there is continual monitoring, it’s particularly easy for interviewers to cheat in telephone surveys—from failing to probe, to interviewing unqualified respondents, to fabricating an item response, and even to fabricating whole interviews. Kiecker and Nelson (1996) hired 33 survey research companies to do eight interviews each, ostensibly as “mop-up” for a larger national market survey. The eight respondents were plants—graduate students of drama, for whom this must have been quite a gig—and were the same eight for each of the surveys. Of the 33 interviewers studied, 10 fabricated an entire interview, 32 fabricated at least one item response, and all 33 failed to record responses verbatim.

The technology of telephone interviewing has become very sophisticated. **Computer-assisted telephone interviewing** (CATI) makes it harder to do things like ask questions out of order, but a determined cheater on your interviewing team can do a lot of damage. The good news is that once you eliminate cheating (with monitoring), the main thing left that can go wrong is inconsistency in the way interviewers ask questions. Unstructured and structured interviews each have their own advantages, but for structured interviews to yield reliable results, they have to be really, really structured. That is, the questions have to be read verbatim so that every respondent is exposed to the same stimulus.

Repeated verbatim readings of questions is boring to do and boring to listen

to. When respondents (inevitably) get restless, it's tempting to vary the wording to make the interview process seem less mechanical. This turns out to be a bigger problem in face-to-face interviews (where interviewers are generally working alone, without any monitoring) than in telephone interviews. Presser and Zhao (1992) monitored 40 trained telephone interviewers at the Maryland Survey Research Center. For the 5,619 questions monitored, interviewers read the questions exactly as worded on the survey 91% of the time. Training works.

Still, no matter how much you train interviewers . . . Johnstone et al. (1992) studied 48 telephone interviews done entirely by women and found that female respondents elicited more sympathy, while male respondents elicited more joking. Men, say Johnstone et al., may be less comfortable than women are with being interviewed by women and wind up trying to subvert the interview by turning it into teasing or banter.

Sampling for telephone surveys is also aided by computer. There are companies that sell telephone numbers for surveys. The numbers are chosen to represent businesses or residences and to represent the varying saturation of phone service in different calling areas.

Even the best sample of phone numbers, though, may not be enough to keep you out of trouble. During the 1984 U.S. presidential election, Ronald Reagan's tracking poll used a list of registered voters, Republicans and Democrats alike. The poll showed Reagan comfortably ahead of his rival, Walter Mondale, except on Friday nights. Registered Republicans, it turned out, being wealthier than their counterparts among Democrats, were out Friday nights more than Democrats were, and simply weren't available to answer the phone (Begley et al. 1992:38).

When to Use What

There is no perfect data-collection method. However, mailed or dropped-off questionnaires are preferable to personal interviews when three conditions are met: (1) You are dealing with literate respondents; (2) You are confident of getting a high response rate (at least 70%); and (3) The questions you want to ask do not require a face-to-face interview or the use of visual aids such as cue cards, charts, and the like. Under these circumstances, you get much more information for your time and money than from the other methods of questionnaire administration.

When you really need complete interviews—answers to all or nearly all the questions in a particular survey—then face-to-face interviews, whether assisted by computer or not, are the way to go. Caserta et al. (1985) inter-

viewed recently bereaved respondents about adjustment to widowhood. They interviewed 192 respondents—104 in person, at home, and 88 by mailed questionnaire. Both groups got identical questions.

On average, 82% of those interviewed at home 3–4 weeks after losing their husband or wife answered any given question. Just 68% of those who responded to the mailed questionnaire answered any given question. As Caserta et al. explain, the physical presence of the interviewer helped establish the rapport needed for asking sensitive and personal questions about the painful experience of bereavement (*ibid.*:640).

If you are working in a highly industrialized country, and if a very high proportion (at least 80%) of the population you are studying has their own telephones, then consider doing a phone survey whenever a self-administered questionnaire would otherwise be appropriate.

If you are working alone or in places where the mails and the phone system are inefficient for data collection, the drop-and-collect technique is a good alternative (see above, page 258).

Finally, there is no rule against using more than one type of interview. Mauritius, an island nation in the Indian Ocean, is an ethnically complex society. Chinese, Creoles, Franco-Mauritians, Hindus, Muslims, and other groups make up a population of about a million. Ari Nave (1997) was interested in how Mauritians maintain their ethnic group boundaries, particularly through their choices of whom to marry. A government office on Mauritius maintains a list of all people over 18 on Mauritius, so it was relatively easy for Nave to get a random sample of the population.

Contacting the sample was another matter. Nave got back just 347 out 930 mailed questionnaires, but he was able to interview another 296 by telephone and face to face, for a total of 643, or 69% of his original sample—a respectable completion rate.

Using Interviewers

There are several advantages to using multiple interviewers in survey research. The most obvious is that you can increase the size of the sample. Multiple interviewers, however, introduce several disadvantages, and whatever problems are associated with interviewer bias are increased with more than one interviewer.

Just as important, multiple interviewers increase the cost of survey research. If you can collect 100 interviews yourself and maintain careful quality control in your interview technique, then hiring one more interviewer would probably not improve your research by enough to warrant both spend-

ing the extra money and worrying about quality control. Recall that for estimating population proportions or means, you have to quadruple the sample size to halve the sampling error. If you can't afford to hire three more interviewers (beside yourself), and to train them carefully so that they at least introduce the *same* bias to every interview as you do, you're better off running the survey yourself and saving the money for other things.

This only goes for surveys in which you interview a random sample of respondents in order to estimate a population parameter. If you are studying the experiences of a group of people, or are after cultural data (as in "How are things usually done around here?"), then getting more interviews is better than getting fewer, whether you collect the data yourself or have them collected by others.

Training Interviewers

If you hire interviewers, be sure to train them—and monitor them throughout the research. A colleague used a doctoral student as an interviewer in a project in Atlanta. The senior researcher trained the student but listened to the interview tapes that came in. At one point, the interviewer asked a respondent: "How many years of education do you have?" "Four," said the respondent. "Oh," said the student researcher, "you mean you have four years of education?" "No," said the informant, bristling and insulted, "I've had four years of education beyond high school." The informant was affluent; the interview was conducted in his upper-middle-class house; he had already told the interviewer that he was in a high-tech occupation. So monitor interviewers.

If you hire a *team* of interviewers, you have one extra chore besides monitoring their work. You need to get them to act as a team. Be sure, for example, that they all use the same probes to the various questions on the interview schedule. Especially with open-ended questions, be sure to do random spot checks, *during the survey*, of how interviewers are coding the answers they get. The act of spot-checking keeps coders alert. When you find discrepancies in the way interviewers code responses, bring the group together and discuss the problem openly.

Narratives are coded after the interview. If you use a team of coders, be sure to train them together and get their interrater reliability coefficient up to at least .70. In other words, make sure that your interviewers use the same theme tags to code each piece of text. For details on how to do this, see the section on Cohen's Kappa in chapter 17.

Billiet and Loosveldt (1988) found that asking interviewers to tape all their interviews produces a higher response rate, particularly to sensitive questions about things like sexual behavior. Apparently, when interviewers know that

their work can be scrutinized (from the tapes), they probe more and get informants to open up more.

Carey et al. (1996) studied the beliefs of 51 newly arrived Vietnamese refugees in upstate New York about tuberculosis. The interviews consisted of 32 open-ended questions on beliefs about symptoms, prevention, treatment, and the social consequences of having TB. The two interviewers in this study were bilingual refugees who participated in a 3-day workshop to build their interview skills. They were told about the rationale for open-ended questions and about techniques for getting respondents to open up and provide full answers to the questions. The training included a *written manual* (this is very important) to which the interviewers could refer during the actual study. After the workshop, the trainees did 12 practice interviews with Vietnamese adults who were not in the study.

William Axinn ran the Tamang Family Research Project, a comparative study of villages in Nepal (Axinn et al. 1991). Axinn and his coworkers trained a group of interviewers using the *Interviewer's Manual* from the Survey Research Center at the University of Michigan (University of Michigan 1976). That manual contains the distilled wisdom of hundreds of interviewer training exercises in the United States, and Axinn found the manual useful in training Nepalese interviewers, too.

Axinn recruited 32 potential interviewers. After a week of training (5 days at 8 hours a day, and 2 days of supervised field practice), the 16 best interviewers were selected, 10 men and 6 women. The researchers hired more interviewers than they needed and after 3 months, four of the interviewers were fired. "The firing of interviewers who clearly failed to follow protocols," said Axinn et al., "had a considerable positive effect on the morale of interviewers who had worked hard to follow our rules" (1991:200). No one has accused Axinn of overstatement.

Whom to Hire

In general, when hiring interviewers, look for professional interviewers first. Next, look for people who are mature enough to accept the need for rigorous training and who can work as part of a team. If need be, look for interviewers who can handle the possibility of going into some rough neighborhoods and who can answer the many questions that respondents will come up with in the course of the survey.

If you are running a survey based on personal interviews in a developing country, consider hiring college students, and even college graduates, in the social sciences. "Social sciences," by the way, does not mean the humanities. In Peru, Donald Warwick and Charles Lininger found that "some students

from the humanities . . . were reluctant to accept the ‘rigidities’ of survey interviewing.” Those students felt that “As educated individuals, they should be allowed to administer the questionnaire as they saw fit in each situation” (Warwick and Lininger 1975:222).

I would not use anyone who had that kind of attitude as an interviewer. But undergraduate social science students in the developing world may have real research experience since most of them aren’t going on for graduate training. Students who are experienced interviewers have a lot to contribute to the design and content of questionnaires. Remember, you are dealing with colleagues who will be justly resentful if you treat them merely as employees of your study. By the same token, college students in developing nations—particularly those in public universities—are likely to be members of the elite who may find it tough to establish rapport with peasant farmers or the urban poor (Hursh-César and Roy 1976:308).

Make It Easy for Interviewers to Do Their Job

If you use interviewers, be sure to make the questionnaire booklet easy to use. Leave enough space for interviewers to write in the answers to open-ended questions—but not too much space. Big spaces are an invitation to some interviewers to develop needlessly long answers (Warwick and Lininger 1975:152).

Also, use two different type faces for questions and answers; put instructions to interviewers in capital letters and questions for respondents in normal type. Figure 10.2 is an example:

5. INTERVIEWER: CHECK ONE OF THE FOLLOWING
- R HAS LIVED IN CHICAGO MORE THAN FIVE YEARS.
SKIP TO QUESTION 7.
 - R HAS LIVED IN CHICAGO LESS THAN FIVE YEARS.
ASK QUESTION 6 AND CONTINUE WITH QUESTION 7.
6. Could you tell me where you were living five years ago?
7. Where were you born?

Figure 10.2. Using two different type faces in a survey instrument.

SOURCE: Adapted from D. P. Warwick and C. A. Lininger, *The Sample Survey: Theory and Practice*, p. 153. © 1975, McGraw-Hill; rights reverted to authors.

Closed Vs. Open-Ended Questions

The most often-asked question about survey research is whether **fixed-choice** (also called **closed-ended**) or **open-ended** items are better. The answer

is that the two formats produce different kinds of data, and it's your call when to use what. One obvious disadvantage of fixed-choice questions is that people naturally focus on the choices they have. If they'd like to offer a response other than those in front of them, they won't do it, even if they can (Krosnick 1999:544). Schuman and Presser (1981:89) asked a sample of people this question: "Please look at this card and tell me which thing you would most prefer in a job." The card had five items listed: (1) high income, (2) no danger of being fired, (3) working hours are short—lots of free time, (4) chances for advancement, and (5) the work is important and gives a feeling of accomplishment. Then they asked a different sample the open-ended question: "What would you most prefer in a job?" About 17% of the respondents to the fixed-choice question chose "chances for advancement," and over 59% chose "important work." Under 2% of the respondents who were asked the open-ended question mentioned "chances for advancement," and just 21% said anything about "important" or "challenging" or "fulfilling" work.

When the questions get really threatening, fixed-choice questions are generally not a good idea. Masturbation, alcohol consumption, and drug use are reported with 50%–100% greater frequency in response to open-ended questions (Bradburn 1983:299). Apparently, people are least threatened when they can offer their own answers to open-ended questions on a self-administered questionnaire, rather than being forced to choose among a set of fixed alternatives (e.g., once a month, once a week, once a day, several times a day), and are most threatened by a face-to-face interviewer (Blair et al. 1977).

On the other hand, Ivis et al. (1997) found that at least one pretty embarrassing question was better asked in a fixed-choice format—and over the phone, at that. People in their survey were asked: "How often in the last 12 months have you had five or more drinks on one occasion?" Then, later in the interview, they were asked the same question, but were given nine fixed choices: (1) every day; (2) about once every other day; . . . (9) never in the last year. The fixed-choice format produced significantly more positive responses. The anonymity of telephone surveys provides a certain comfort level where people feel free to open up on sensitive topics. And notice that the anonymity of telephone surveys lets the *interviewer*, as well as the respondent, off the hook. You can ask people things you might be squeamish about if the interview were face to face, and respondents feel that they can divulge very personal matters to disembodied voices on the phone.

Overall, since closed-ended items are so efficient, most survey researchers prefer them to open-ended questions and use them whenever possible. There is no rule, however, that prevents you from mixing question types. Many survey researchers use the open-ended format for really intimidating questions and the fixed-choice format for everything else, even on the phone. Even if there are no intimidating questions in a survey, it's a good idea to stick in a few

open-ended items. The open-ended questions break the monotony for the respondent, as do tasks that require referring to visual aids (like a graph).

The responses to fixed-choice questions are unambiguous for purposes of analysis. Be sure to take full advantage of this and **precode fixed-choice** items on a questionnaire. Put the codes right on the instrument so that typing the data into the computer is as easy (and as error free) as possible.

It's worth repeating that when you do computer-assisted interviews in the field (CAPI, CASI, audio CASI) you cut down on data entry error. The fewer times you have to touch data, the fewer opportunities there are to stick errors in them. I particularly like the fact that we can combine fixed-choice and open-ended questions on a hand-held computer for fieldwork. (For more on this technology, see Gravlee 2002a. For more on the efficacy of various survey formats—self-administered, face-to-face, telephone—see Wentland and Smith 1993. And for more on the virtues of open-ended questions when you're studying sensitive issues, see Schaeffer 2000 and Levy-Storms et al. 2002.)

Question Wording and Format

There are some well-understood rules that all survey researchers follow in constructing questionnaire items. Here are 15 of them.

1. *Be unambiguous.* If respondents can interpret a question differently from the meaning you have in mind, they will. In my view, this is the source of most response error in fixed-choice questionnaires.

The problem is not easy to solve. A simple question like “How often do you visit a doctor?” can be very ambiguous. Are acupuncturists, chiropractors, chiropractists, and public clinics all doctors? If you think they are, you'd better tell people that, or you leave it up to them to decide. In some parts of the southwestern United States, people may be visiting native curers and herbalists. Are those practitioners doctors? In Mexico, many community clinics are staffed by nurses. Does “going to the doctor” include a visit to one of those clinics?

Here's how Cannell et al. (1989) recommend asking about doctor visits in the last year:

Have you been a patient in the hospital overnight in the past 12 months since July 1st 1987?

(Not counting when you were in a hospital overnight.) During the past 12 months since July 1st, 1987, how many times did you actually see any medical doctor about your own health?

During the past 12 months since July 1st 1987, were there any times when you didn't actually see the doctor but saw a nurse or other medical assistant working for the doctor?

During the past 12 months since July 1st 1987, did you get any medical advice, prescriptions, or results of tests over the telephone from a medical doctor, nurse, or medical assistant working for a doctor? (Cannell et al. 1989, appendix A:1, cited in Schaeffer and Presser 2003:71)

If you ask: "How long have you lived in Mexico City?" does "Mexico City" include the 20 million people who live in the urban sprawl, or just the eight million who live in the Federal District? And how "near" is "near Nairobi"?

Words like "lunch," "community," "people," and hundreds of other innocent lexical items have lurking ambiguities associated with them, and phrases like "family planning" will cause all kinds of mischief. Half the respondents in the 1985 General Social Survey were asked if they agreed that there was too little spending for "assistance to the poor," while half were asked if there was too little spending for "welfare." A whopping 65% agreed with the first wording; just 19% agreed with the second (Smith 1987:77).

Even the word "you," as Payne pointed out (1951), can be ambiguous. Ask a nurse at the clinic "How many patients did you see last week?" and you might get a response like: "Who do you mean, me or the clinic?" If the nurse is filling out a self-administered questionnaire, she'll have to decide for herself what you had in mind. Maybe she'll get it right; maybe she won't.

2. *Use a vocabulary that your respondents understand, but don't be condescending.*
This is a difficult balance to achieve. If you're studying a narrow population (sugar cane cutters, midwives, leather workers), then proper ethnography and pretesting with a few knowledgeable informants will help ensure appropriate wording of questions.

But if you are studying a more general population, even in a small town of just 3,000 people, then things are very different. Some respondents will require a low-level vocabulary; others will find that vocabulary insulting. This is one of the reasons often cited for doing personal interviews: You want the opportunity to phrase your questions differently for different segments of the population. Realize, however, that this poses risks in terms of reliability of response data.

3. *Remember that respondents must know enough to respond to your questions.*
You'd be surprised at how often questionnaires are distributed to people who are totally unequipped to answer them. I get questionnaires in the mail and by e-mail all the time, asking for information I simply don't have.

Most people can't recall with any acceptable accuracy how long they spent in the hospital last year, how many miles they drive each week, or how much they've cut back on their use of air-conditioning. They *can* recall whether they own a television, have *ever* been to Cairo, or voted in the recent elections. And they can tell you whether they *think* they got a fair price for the land they vacated when the dam was built or *believe* the local member of parliament is doing a better job than her predecessor at giving equal time to rich people and poor people who come to her with complaints.

4. Make sure there's a *clear purpose for every question* you ask in a survey. When I say "clear purpose," I mean clear to respondents, not just to you. And once you're on a topic, stay on it and finish it. Respondents can get frustrated, confused, and annoyed at the tactic of switching topics and then coming back to a topic that they've already dealt with on a questionnaire. Some researchers do exactly this just to ask the same question in more than one way and to check respondent reliability. This underestimates the intelligence of respondents and is asking for trouble—I have known respondents to sabotage questionnaires that they found insulting to their intelligence.

You can (and should) ask questions that are related to one another at different places in a questionnaire, so long as each question makes sense in terms of its placement in the overall instrument. For example, if you are interviewing labor migrants, you'll probably want to get a labor history—by asking where the respondent has worked during the past few years. Later, in a section on family economics, you might ask whether a respondent has ever sent remittances and from where.

As you move from one topic to another, put in a transition paragraph that makes each shift logical to the respondent. For example, you might say: "Now that we have learned something about the kinds of food you like, we'd like to know about. . . ." The exact wording of these transition paragraphs should be varied throughout a questionnaire.

5. Pay careful attention to **contingencies** and **filter questions**. Many question topics contain several contingencies. Suppose you ask someone if they are married. If they answer "no," then you probably want to ask whether they've ever been married. You may want to know whether they have children, irrespective of whether they are married or have ever been married. You may want to know what people think is the ideal family size, irrespective of whether they've been married, plan to be married, have children, or plan to have children.

You can see that the contingencies can get very complex. The best way to ensure that all contingencies are accounted for is to build a **contingency flow chart** like that shown in figure 10.3 (Sirken 1972; Sudman and Bradburn 1982).

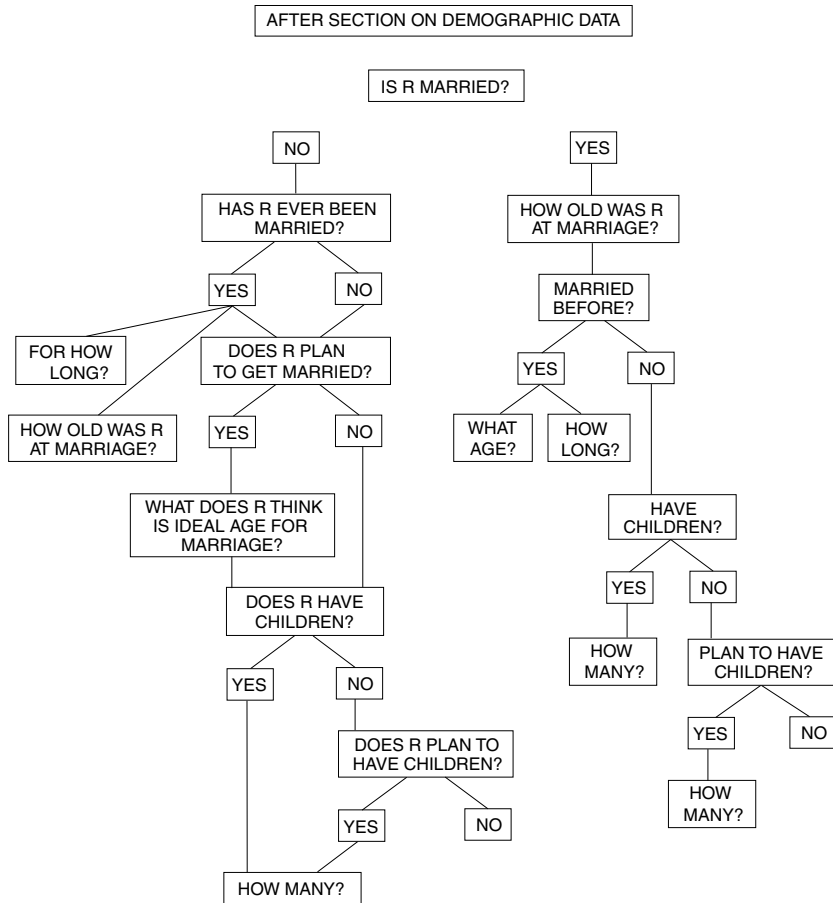


Figure 10.3. Flow chart of filter questions for part of a questionnaire.

6. *Use clear scales.* There are some commonly used scales in survey research—things like: Excellent-Good-Fair-Poor; Approve-Disapprove; Oppose-Favor; For-Against; Good-Bad; Agree-Disagree; Better-Worse-About the Same; etc. Just because these are well known, however, does not mean that they are clear and unambiguous to respondents.

To cut down on the ambiguities associated with these kinds of scales, explain the meaning of each potentially ambiguous scale when you introduce it. With self-administered questionnaires, use 5 scale points rather than 3, if you can. For example, use Strongly Approve, Approve, Neutral, Disapprove,

Strongly Disapprove, rather than Approve, Neutral, Disapprove. This will give people the opportunity to make finer-grained choices. If your sample is large enough, you can distinguish during analysis among respondents who answer, say, “strongly approve” vs. “approve” on some item.

For smaller samples, you’ll have to aggregate the data into three categories for analysis. Self-administered questionnaires allow the use of 7-point scales, like the semantic differential scale shown in figure 10.4, and even longer scales. Telephone interviews usually require 3-point scales.

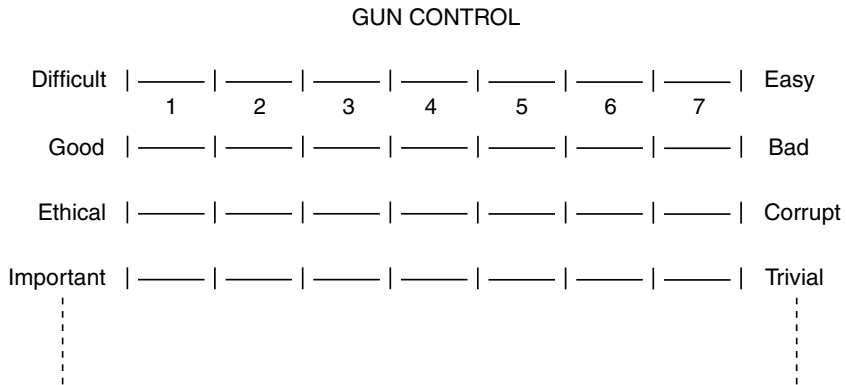


Figure 10.4. A 7-point semantic differential scale.

Notice that the semantic differential scale in figure 10.4 has word anchors at both ends and numbers in the middle, not words. In this kind of scale, we want to let people interpret the dimension indicated by the anchors. In typical rating scales (you know, the 3- and 5-point scales you see in questionnaires), we want to remove ambiguity, so we label all the points in words—like Strongly Agree, Agree, Neutral, Disagree, Strongly Disagree (Peters and McCormick 1966. Much more on how to construct scales in chapter 12.)

7. Try to *package questions in self-administered questionnaires*, as shown earlier in figure 10.1. This is a way to get a lot of data quickly and easily, and, if done properly, it will prevent respondents from getting bored with a survey. For example, you might say “Please indicate how close you feel to each of the persons on this chart” and provide the respondent with a list of relatives (mother, father, sister, brother, etc.) and a scale (very close, close, neutral, distant, very distant, etc.).

Be sure to make scales unambiguous. If you are asking how often people think they do something, don’t say “regularly” when you mean “more than

once a month,” and limit the list of activities to no more than seven. Then introduce a question with a totally different format, to break up the monotony and to keep the respondent interested.

Packaging is best done in self-administered questionnaires. If you use these kinds of lists in a face-to-face interview, you’ll have to repeat the scale for at least the first three items or activities you name, or until the respondent gets the pattern down. This can get very tiring for both interviewers and respondents.

8. If you want respondents to check just one response, then be sure to *make the possible responses to a question exhaustive and mutually exclusive*. This may mean including a “don’t know” option.

Here is an example (taken from a questionnaire I received) of what *not* to do:

How do you perceive communication between your department and other departments in the university? (check one)

- | | |
|-----------------------------------|-------|
| There is much communication | _____ |
| There is sufficient communication | _____ |
| There is little communication | _____ |
| There is no communication | _____ |
| No basis for perception | _____ |

The “no basis for perception” response took care of making the item exhaustive. You can always make questionnaire items like this one exhaustive by giving respondents the option of saying some variant of “don’t know”—like “no basis for perception.” Some researchers feel that this just gives respondents a lazy way out—that people need to be made to work a bit. If there is a good chance that some of your respondents really won’t have the information you ask for, then I think the “don’t know” option is too important to leave out. In consumer preference surveys, though, where you actually give someone a taste of a cracker and ask them to tell you if they like it, the “don’t know” option is a bad idea.

The problem for me on this item was that I wanted to check both “little communication” and “sufficient communication.” For me, at least, these two categories were not mutually exclusive—I didn’t think there was a lot of communication, and I wasn’t at all bothered by that—but the author of the survey asked me to “check one.” (For more on the “don’t know” option in surveys, see Lam et al. 2002.)

9. *Keep threatening questions short.* Many questions require a preamble in order to set up a time frame or otherwise make clear what you are asking an informant to think about in answering a question. For example:

These next questions are about what your children are doing these days. You said that you have two daughters and a son. Where is your older daughter living? Your younger daughter? Your son?

These next questions are about your travels to sell *huipiles* in the last year. Since October 2005, how many times have you been to Chichicastenango to sell your *huipiles*?

Questions that are likely to intimidate respondents should have long preambles to lessen the intimidation effect. The questions themselves, however, should contain as few words as possible.

10. *Always provide alternatives*, if appropriate. Suppose people are being asked to move off their land to make way for a new highway. The government offers to compensate people for the land, but people are suspicious that the government won't evaluate fairly how much compensation landowners are entitled to. If you take a survey and ask "Should the government offer people compensation for their land?" respondents can answer yes or no for very different reasons. Instead, let people check whether they agree or disagree with a set of alternatives, like: "The government should offer people compensation for their land" and "An independent board should determine how much people get for their land."
11. *Avoid loaded questions.* Any question that begins "Don't you agree that . . ." is a **loaded question**. Sheatsley (1983) points out, however, that asking loaded questions is a technique you can use to your advantage, on occasion, just as leading or baiting informants can be used in unstructured interviewing. A famous example comes from Kinsey's landmark study of sexual behavior of American men (Kinsey et al. 1948). Kinsey asked men "How old were you the first time you masturbated?" This made respondents feel that the interviewer already *knew* about the fact of masturbation and was only in search of additional information.
12. *Don't use double-barreled questions.* Here is one I found on a questionnaire: "When did you leave home and go to work on your own for the first time?" There is no reason to assume, of course, that someone had to leave home in order to go to work, or that they necessarily went to work if they left home.

Here is another bad question:

Please indicate if you agree or disagree with the following statement:
 Marijuana is no more harmful than tobacco or alcohol, so the personal use of marijuana should be legal.

Suppose a respondent agrees (or disagrees) with the first part of the statement—the assertion that marijuana is no more harmful than tobacco or alcohol. He or she may agree or disagree with the second part of the statement. If respondents answer "yes" or "no," how do you know if they are indicating

agreement with both parts of it or just one part? Which part? How can you tell? You can't. That's why it's a bad question.

13. *Don't put false premises into questions.* I once formulated the following question for a survey in Greece: "Is it better for a woman to have a house and cash as a dowry, or for her to have an education and a job that she can bring to the marriage?" This question was based on a lot of ethnographic work in a community, during which I learned that many families were sinking their resources into getting women educated and into jobs and offering this to eligible bachelors as a substitute for traditional material dowries.

My question, however, was based on the false premise that all families respected the custom of dowry. The question did not allow respondents to state a third alternative—namely, that they didn't think dowry was a custom that ought to be maintained in any form, traditional or modern. In fact, many families were deciding to reject the dowry custom altogether—something that I missed for some time because I failed to pretest the item (see Pretesting, below).

14. *Don't take emotional stands in the wording of questions.* Here's an example of the sort of question you see on surveys all the time—and that you should never ask: "Should the legislature raise the drinking age to 21 in order to reduce the carnage among teens on our highways?" Another example of a bad question is: "Don't you agree with the President when he says . . . ?"
15. When asking for opinions *on controversial issues, specify the referent situation* as much as possible. Instead of asking: "Do you approve of abortion?" ask: "Under what conditions do you approve of abortion?" Then give the respondent an exhaustive list of circumstances as possible to check. If the circumstances are not exclusive (rape and incest are not necessarily exclusive, for example), then let respondents check as many circumstances as they think appropriate.

Translation and Back Translation

All the tips given here about writing good survey questions continue to apply when you are working in another culture. They are just a lot more difficult to implement because you have to deal with phrasing questions properly in another language as well. The best way to deal with this is through back translation (Brislin 1970; Werner and Campbell 1970).

Back translation is the standard method for adapting social and psychological measurement scales. Look up "back translation" in a database like PsycINFO and you'll find over a hundred examples. First, write any questionnaire in *your* native language, paying attention to all the lessons of this chapter. Then have the questionnaire translated by a bilingual person who is a native speaker of the language you are working in. Work closely with the translator, so

that she or he can fully understand the subtleties you want to convey in your questionnaire items.

Next, ask another bilingual person, who is a native speaker of *your* language, to translate the questionnaire back into that language. This back translation should be almost identical to the original questionnaire you wrote. If it isn't, then something was lost in one of the two translations. You'd better find out which one it was and correct the problem.

Beck and Gable (2000) developed a scale for screening postpartum women for depression and then translated the scale into Spanish (Beck and Gable 2003). One item on the original scale was "I felt like my emotions were on a roller coaster." The first translator offered two options for this: "Sentí un sube y baja emocional" and "Sentí un desequilibrio emocional." The second translator translated these as "I felt like my emotions were up and down" and "I felt emotional instability" (ibid.:69). Not exactly the same feeling as "I felt like my emotions were on a roller coaster," but close. Do you go with one of the two Spanish translations offered? Which one? Or do you keep looking for something better in Spanish? The answer is that you sit down with both translators, talk it through, and come to a consensus.

You can also use back translation to check the content of open-ended interviews, but be warned: This is tough work. Daniel Reboussin (1995) interviewed Diola women who had come from southwestern Senegal to Dakar in search of work. All the women used French at work, but they preferred Diola for interviews. Reboussin, who speaks French, spoke very little Diola, so he worked with an interpreter—a man named Antoine Badji—to develop an interview schedule in French, which Badji translated into Diola.

During the interviews, Badji asked questions in Diola and Reboussin audio-taped the responses. After each interview, Badji translated each tape (orally) into French. Reboussin transcribed the French translations, translating into English as he went. Then he read his English transcriptions *back* to Badji, translating (orally) into French as he read. That way, Badji could confirm or disconfirm Reboussin's understanding of Badji's French rendering of the tapes.

As I said, this was tough work. It took Reboussin and Badji 17 weeks to conduct 30 interviews and get them all down into English.

The Response Rate Problem

Mailed questionnaires can be very, very effective, but there is one problem with them that all survey researchers watch for: getting enough of them back. In 1936, the *Literary Digest* sent out 10 million straw poll ballots in an attempt

to predict the winner of the presidential election. They got back 2.3 million ballots and predicted Alf Landon over Franklin Delano Roosevelt in a landslide. Roosevelt got 61% of the vote.

Now, you'd think that 2.3 million ballots would be enough for anyone, but two things caused the *Digest* debacle. First, they selected their sample from automobile registries and telephone books. In 1936, this favored richer people who tend to be Republican. Second, the 2.3 million ballots were only 23% of the 10 million sent out. The low response rate biased the results in favor of the Republican challenger since those who didn't respond tended to be poorer and less inclined to participate in surveys (Squire 1988).

How to Adjust for Nonresponse

Skip to 1991. The American Anthropological Association sent questionnaires to a sample of 1,229 members. The sample was stratified into several cohorts who had received their Ph.D. degrees beginning in 1971–1972 and ending in 1989–1990. The 1989–1990 cohort comprised 306 then-recent Ph.D.s. The idea was to find out what kinds of jobs those anthropologists had.

The AAA got back 840 completed questionnaires, or 68% of the 1,229, and the results of the survey were reported in the *Anthropology Newsletter* in May 1991. The response rate is not high for this kind of survey, where respondents are being asked for information from their own professional organization. (The U.S. Office of Management and Budget demands a minimum 75% response rate from survey contract researchers [Fowler 1984:48] and, as we saw earlier, this is not an excessive demand.)

Now, 41% of those responding from the 1989–1990 cohort said they had academic jobs. The *Anthropology Newsletter* didn't report the response rate by cohort, but suppose that 68% of the 1989–1990 cohort—the same percentage as applies to the overall survey—sent back their questionnaires. That's 208 out of 306 responses. The 41% who said they had academic jobs would be 85 of the 208 respondents; the other 123 had nonacademic jobs.

Suppose that everyone who didn't respond (32%, or 98 out of 306) got non-academic jobs. (Maybe that's why they didn't bother to respond.) In that case, $98 + 123 = 221$ out of the 306 people in the cohort, or 72%, got nonacademic jobs that year—not the 59% ($100\% - 41\%$) as reported in the survey.

It's unlikely that *all* the nonresponders were in nonacademic jobs. To handle the problem of nonresponse, the AAA might have run down a random grab of 10 of the nonresponders and interviewed them by telephone. Suppose that seven said they had nonacademic jobs. You'll recall from chapter 7 on sampling theory that the formula for determining the 95% confidence limits of a point estimator is:

$$P \text{ (the true proportion)} = 1.96 \left(\sqrt{PQ/n} \right) \quad \text{Formula 10.1}$$

which means that

$$1.96 \left(\sqrt{(.70)(.30)/10} \right) = .28$$

The probable answer for the 10 holdouts is $.70 + .28$. Somewhere between 42% and 98% of the 98 nonresponders from the 1989–1990 cohort probably had nonacademic jobs. We guess that between 41 and 96 of those 98 nonresponders had nonacademic jobs. We can now make a reasonable guess: 123 of the responders plus anywhere from 41 to 96 of the nonresponders had nonacademic jobs, which means that between 164 and 216 of the 306 people in the cohort, or 54% to 71%, probably had nonacademic jobs.

Low response rate can be a disaster. People who are quick to fill out and return mailed questionnaires tend to have higher incomes and consequently tend to be more educated than people who respond later. Any dependent variables that covary with income and education, then, will be seriously distorted if you get back only 50% of your questionnaires. And what's worse, there is no accurate way to measure nonresponse bias. With a lot of nonresponse, all you know is that you've got bias but you don't know how to take it into account.

Improving Response Rates: Dillman's Total Design Method

Fortunately, a lot of research has been done on increasing response rates to mailed questionnaires. Yammarino et al. (1991) reviewed 184 controlled experiments, done between 1940 and 1988, on maximizing the return of mailed questionnaires, and Don Dillman, of the Survey Research Laboratory at Washington State University, has synthesized the research on maximizing return rates and has developed what he calls the "Total Design Method" (TDM) of mail and telephone surveying (Dillman 1978, 1983; Salant and Dillman 1994).

Professional surveys done in the United States, following Dillman's method achieve an average return rate of around 73%, with many surveys reaching an 85%–90% response rate. In Canada and Europe, around 79% of personal interviews are completed, and the response rate for mailed questionnaires is around 75% (Dillman 1978, 1983). Of course, those numbers are for the usual kinds of surveys about consumer behaviors, political attitudes, and so on. What happens when you ask people really threatening questions? In the Neth-

erlands, Nederhof (1985) conducted a mail survey on attitudes toward suicide and achieved a 65% response rate. Pretty impressive. Outside of North America and northern Europe, Jussaume and Yamada (1990) achieved a response rate of 56% in Kobe, Japan, and de Rada (2001) had a response rate of 61% in mostly rural Navarra Province in Spain.

The average response rate for face-to-face interviews in the United States was between 80% and 85% during the 1960s, but fell to less than 70% in the early 1970s (American Statistical Association 1974). It has apparently recovered somewhat as more is learned about how to maximize cooperation by potential respondents.

Willimack et al. (1995), for example, report a refusal rate of 28% for face-to-face interviews in the annual Detroit Area Study conducted by the University of Michigan. They also report that giving people a gift-type ballpoint pen (a small, nonmonetary incentive) *before* starting a face-to-face interview lowered refusal rates and increased the completeness of responses to questions. The bottom line is that, with everything we've learned over the years about how to do mailed surveys, the gap between the response rate to personal interviews and mailed questionnaires is now insignificant, when everything is done right, but the rate of refusals is high in all forms of surveys.

Many scholars are concerned with the response rate problem (Roth and BeVier 1998; Turley 1999; Synodinos and Yamada 2000), but it remains to be seen how much refusals affect the outcomes of surveys—that is, the ability to say accurate things about the population we are studying (Krosnick 1999; Morin 2004). This does not in any way reduce the value of personal interviews, especially for anyone working in developing nations. It does mean, however, that if you are conducting *mailed* survey research in the United States, Canada, Western Europe, Australia, New Zealand, or Japan, you should use Dillman's method.

Steps in Dillman's Method

1. **Professionalism:** Mailed questionnaires must look thoroughly professional. Jaded, hard-bitten, oversurveyed people simply don't respond to amateurish work. Fortunately, with today's word processing, making attractive questionnaire booklets is easy. Use standard-size paper: 8.5" × 11" in the United States and slightly longer, A4 paper, in the rest of the world. Several researchers have found that light green paper produces a higher response rate than white paper (Fox et al. 1988). I wouldn't use any other colors until controlled tests are made.

You must be thinking: "Controlled tests of *paper color*?" Absolutely. It's because social scientists have done their homework on these little things that

a response rate of over 70% is achievable—provided you're willing to spend the time and money it takes to look after all the little things. Read on and you'll see how small-but-important those "little things" are.

2. **Front and back covers:** Don't put any questions on either the front or back covers of the booklet. The front cover should contain a title that provokes the respondent's interest and some kind of eye-catching graphic design. By "provoking interest," I don't mean "threatening." A title like "The Greenville Air Quality Survey" is fine. "Polluted Air Is Killing Us" isn't.

Graphic designs are better than photographs on survey covers. Photos contain an enormous amount of information, and you never know how respondents will interpret the information. If a respondent thinks a photo contains an editorial message (in favor of or against some pet political position), then the survey booklet goes straight into the trash.

The front cover should also have the name and return address of the organization that's conducting the survey.

The back cover should contain a *brief* note thanking the respondent and inviting open-ended comments about the questionnaire. Nothing else.

3. **Question order:** Pay careful attention to question order. Be sure that the first question is directly related to the topic of the study (as determined from the title on the front of the booklet); that it is interesting and easy to answer; and that it is nonthreatening. Once someone starts a questionnaire or an interview, they are very likely to finish it. Introduce threatening questions well into the instrument, but don't cluster them all together.

Put general socioeconomic and demographic questions at the end of a questionnaire. These seemingly innocuous questions are threatening to many respondents who fear being identified (Sudman and Bradburn 1982). Once someone has filled out a questionnaire, they are unlikely to balk at stating their age, income, religion, occupation, etc.

4. **Formatting:** Construct the pages of the questionnaire according to standard conventions. Use upper-case letters for instructions to respondents and mixed upper and lower case for the questions themselves. Never allow a question to break at the end of a page and continue on another page. Mailed surveys have to look good and be easily readable or they get tossed out.

Use plenty of paper; don't make the instrument appear cramped. Line answers up vertically rather than horizontally, if possible. This, for example, is not so good:

Strongly Approve Approve Neutral Disapprove Strongly Disapprove

This is better:

Strongly Approve

Approve

Neutral

Disapprove

Strongly Disapprove

It pays to spend time on the physical format of a questionnaire. The general appearance, the number of pages, the type of introduction, and the amount of white (or green) space—all can affect how people respond, or whether they respond at all (Akkerboom and Schmeets 1998). Once you've gone to the expense of printing up hundreds of survey instruments, you're pretty much stuck with what you've got.

Use lots of open space in building schedules for personal interviews, too. Artificially short, crowded instruments only result in interviewers missing items and possibly in annoying respondents (imagine yourself sitting for 15 minutes in an interview before the interviewer flips the first page of an interview schedule).

5. **Length:** Keep mailed questionnaires down to 10 pages, with no more than 125 questions. Beyond that, response rates drop (Dillman 1978). Herzog and Bachman (1981) recommend splitting questionnaires in half and alternating the order of presentation of the halves to different respondents to test for response effects of questionnaire length.

It is tempting to save printing and mailing costs and to try to get more questions into a few pages by reducing the amount of white space in a self-administered questionnaire. Don't do it. Respondents are never fooled into thinking that a thin-but-crowded questionnaire is anything other than what it seems to be: a long questionnaire that has been forced into fewer pages and is going to be hard to work through.

6. **The cover letter:** A one-page cover letter should explain, in the briefest possible terms, the nature of the study, how the respondent was selected, who should fill out the questionnaire (the respondent or the members of the household), who is funding the survey, and why it is important for the respondent to send back the questionnaire. ("Your response to this questionnaire is very important. We need your response because. . . .")

The one thing that increases response rate more than any other is university sponsorship (Fox et al. 1988). University sponsorship, though, is not enough. If you want a response rate that is not subject to bias, be sure to address the

cover letter directly and personally to the respondent—no “Dear Respondent” allowed—and sign it using a blue ballpoint pen. Ballpoints make an indentation that respondents can see—yes, some people do hold those letters up to the light to check. This marks the letter as having been individually signed. In Japan, Jussaume and Yamada (1990) signed all their letters with an *inkan*, or personal seal, and they wrote the address by hand on the envelope to show that they were serious.

The cover letter must guarantee confidentiality and must explain the presence of an identification number (if there is one) on the questionnaire. Some survey topics are so sensitive that respondents will balk at seeing an identification number on the questionnaire, even if you guarantee anonymity. In this case, Fowler (1984) recommends eliminating the identification number (thus making the questionnaire truly anonymous) and telling the respondents that they simply cannot be identified.

Enclose a printed postcard with the respondent’s name on it and ask the respondent to mail back the postcard *separately* from the questionnaire. Explain that this will notify you that the respondent has sent in the questionnaire so that you won’t have to send the respondent any reminders later on. Fowler found that people hardly ever send back the postcard without also sending back the questionnaire.

7. **Packaging:** Package the questionnaire, cover letter, and reply envelope and postcard in another envelope for mailing to the respondent. Type the respondent’s name and address on the mailing envelope. Avoid mailing labels. Use first-class postage on the mailing envelope and on the reply envelope. Some people respond better to real stamps than to metered—even first-class metered—postage. Hansley (1974) found that using bright commemorative stamps increased response rate.
8. **Inducements:** What about sending people money as an inducement to complete a survey? Mizes et al. (1984) found that offering respondents \$1 to complete and return a questionnaire resulted in significantly increased returns, but offering respondents \$5 did not produce a sufficiently greater return to warrant using this tactic. In 1984, \$5 was close to the value of many respondents’ time for filling out a questionnaire. This makes responding to a survey more like a strictly economic exchange and, as Dillman pointed out, makes it easier for people to turn down (1978:16).

Inflation will surely have taken its toll by this time (sending people \$1 in the mail to answer a survey can’t possibly buy as much response today as it did in the 1980s), but the point is clear: There is a Goldilocks solution to the problem of how much money to send people as an incentive to fill out and return a survey. If you send people too much money or too little, they throw

the survey away. If you send them just the right amount, they are likely to fill out the survey and return it.

Warriner et al. (1996) confirmed this finding. They offered people in Ontario, Canada, \$2, \$5, or \$10 to send back a mailed survey. Factoring in all the costs of follow-up letters, the \$5 incentive produced the most returns for the money.

First-class postage and monetary incentives may seem expensive, but they are cost effective because they increase the response rate. Whenever you think about cutting corners in a survey, remember that all your work in designing a representative sample goes for nothing if your response rate is low. Random samples cease to be representative unless the people in it respond. Also remember that small monetary incentives may be insulting to some people. This is a cultural and socioeconomic class variable that only you can evaluate in your specific research situation. (See Church [1993] for a meta-analysis of the effect of inducements on response rates to mailed surveys.)

9. **Contact and follow-up:** Pay careful attention to contact procedures. Send a letter to each respondent explaining the survey and informing the respondent that a questionnaire will be coming along soon. Send a postcard reminder to all potential respondents a week after sending out the questionnaire. Don't wait until the response rate drops before sending out reminders. Some people hold onto a questionnaire for a while before deciding to fill it out or throw it away. A reminder after 1 week stimulates response among this segment of respondents.

Send a second cover letter and questionnaire to everyone who has not responded 2 weeks later. Finally, 4 weeks later, send another cover letter *and questionnaire*, along with an additional note explaining that you have not yet received the respondent's questionnaire, and stating how important it is that the respondent participate in the study. Heberlein and Baumgartner (1978, 1981) found that sending a second copy of the questionnaire increases response rate 1%–9%. As there does not appear to be any way to predict whether the increase will be 1% or 9%, the best bet is to send the extra questionnaire.

When you send out the second copy of the questionnaire, send the packet by certified mail. House et al. (1977) showed that certified mail made a big difference in return rate for the second follow-up.

Does All This Really Make a Difference?

Thurman et al. (1993) were interested in the attitudes and self-reported behaviors of people who admit to drunken driving. Using Dillman's TDM,

they sent out questionnaires to a national sample of 1,310 and got back 765, or 58%. Not bad for a first pass, since you can generally expect about 25% to 30% from the first wave.

Unfortunately, for lack of time and money, Thurman et al. couldn't follow through with all the extra mailings. Of the 765 respondents, 237 said they were nondrinkers. This left 525 eligible questionnaires for analysis.

Of the 525 respondents who said they were consumers of alcohol, 133 admitted driving while drunk in the past year. Those 133 respondents provided data of intrinsic interest, but the 765 people who responded from the nationally representative sample of 1,310 may be a biased sample on which to base any generalizations. I say "may be" a biased sample because there is no way to tell. And that's the problem.

The bottom line: The last interview you get in any survey—whether you're sending out questionnaires, doing a phone survey, or contacting respondents for face-to-face interviews—is always the most costly and it's almost always worth it. If you really care about representative data, you won't think of all the chasing around you have to do for the last interviews in a set as a nuisance but as a necessary expense of data collection. And you'll prepare for it in advance by establishing a realistic budget of both time and money. (For recent examples of mailed surveys using the TDM, see Gore-Felton et al. [2002] and Filip et al. [2004].)

Pretesting and Learning from Mistakes

There is no way to emphasize sufficiently the importance of **pretesting** any survey instrument. No matter how much you do to prepare a culturally appropriate questionnaire, it is absolutely guaranteed that you will have forgotten something important or that you will have poorly worded one or more vital elements. These glitches can only be identified by pretesting.

If you are building a self-administered questionnaire, bring in at least 6 to 10 pretest respondents and sit with them as they fill out the entire instrument. Encourage them to ask questions about each item. Your pretest respondents will make you painfully aware of just how much you took for granted, no matter how much ethnographic research you did or how many focus groups you ran before making up a questionnaire.

For face-to-face interviews, do your pretesting under the conditions you will experience when the survey is underway for real. If respondents are going to come to your office, then pretest the instrument in your office. If you are going to respondents' homes, then go to their homes for the pretest.

Use the **thinkaloud** method (also called **cognitive testing**) on at least a few

pretest respondents, no matter whether you're doing face-to-face interviews, CAPI, mail surveys, etc. In this method, people think out loud as they decide on how to answer each question in a survey. There are three alternative outcomes with the thinkaloud technique: (1) People understand the question just as you intended them to; (2) People understand the question very well, but not the way you intended them to; and (3) People don't understand the question at all. Edwards et al. (2005) used this method to pretest a 28-question survey on the use of condoms by women sex workers in Mombassa, Kenya. The result was a survey with culturally appropriate vocabulary for various types of sex clients. (For more on the thinkaloud method, see DeMaio and Rothgeb 1996.)

Never use any of the respondents in a pretest for the main survey. If you are working in a small community, where each respondent is precious (and you don't want to use up any of them on a pretest), take the survey instrument to another community and pretest it there. This will also prevent the pretest respondents in a small community from gossiping about the survey before it actually gets underway. A "small community," by the way, can be "the 27 students from Taiwan at your university" or all the residents of an Indonesian rice-farming village.

If you have a team of face-to-face interviewers, make sure they all take part in the pretest—and be sure to do some of the pretesting yourself. After the pretests, bring the interviewers together for a discussion on how to improve the survey instrument. Ask them if people found some questions hard to answer—or even refused to answer. Ask them if they would change the wording of any of the questions. Check all this yourself by watching a couple of interviews done by people on your team and note when informants ask questions and how the interviewers respond. That way, you can train interviewers to respond in the same way to questions from informants.

As you conduct the actual survey, ask respondents to tell you what they think of the study and of the interview they've just been through. At the end of the study, bring all the interviewers back together for an evaluation of the project. If it is wise to learn from your mistakes, then the first thing you've got to do is find out what the mistakes are. If you give them a chance, your respondents and your interviewers will tell you.

Cross-Sectional and Longitudinal Surveys

Most surveys are **cross-sectional**. The idea is to measure some variables at a single time. Of course, people's attitudes and reported behaviors change over time, and you never know if a single sample is truly representative of the

population. Many surveys are conducted again and again to monitor changes and to ensure against picking a bad sample. Multiple cross-sectional polls use a **longitudinal design**. The daily—even hourly—tracking polls in U.S. presidential elections are an extreme example, but in many industrialized countries some questions have been asked of representative samples for many years.

The Gallup Poll, for example, has been asking Americans to list “the most important problem facing this country today” for about 60 years. The data track the concerns of Americans about unemployment, the quality of education, drugs, street crime, the federal deficit, taxes, health care costs, poverty, racism, AIDS, abortion. . . . There are not many surprises in the data. People in the United States are more worried about the economy in recessions, less worried when the economy is clicking along. Only about 10% said that the war in Iraq was the most important problem in America in March 2004, a year after the invasion, but the number jumped to 26% in April and stayed there for a year. The data from the Gallup Poll, and others like it, are important because they were collected with the same instrument. People were asked the same question again and again over the years. After several generations of effort, longitudinal survey data have become a treasured resource in the highly industrialized nations.

Panel Studies

Multiple cross-sectional surveys have their own problems. If the results from two successive samples are very different, you don’t know if it’s because people’s attitudes or reported behaviors have changed, or the two samples are very different, or both. To deal with this problem, survey researchers use the powerful **panel design**. In a panel study, you interview the same people again and again. Panel studies are like true experiments: Randomly selected participants are tracked for their exposure or lack of exposure to a series of interventions in the real world.

The Panel Study on Income Dynamics, for example, has tracked about 8,000 American families since 1969. Some families have been tracked for 36 years now, and new families are added as others drop out of the panel (families are interviewed every other year). Among many other things, the data track the effect of living in different kinds of neighborhoods on the social and economic mobility of men and women, of Whites and African Americans, of old people and young. The PSID is run by the Institute for Social Research at the University of Michigan, which also makes the data available to the public at <http://psidonline.isr.umich.edu/>. Hundreds of papers have been written from the data of this study. (To find papers based on the data from this panel study, look up PSID as a keyword in the electronic databases at your library.)

The Wisconsin Longitudinal Survey is another treasure. In 1957, the University of Wisconsin surveyed all 31,000 high school seniors in the state about their educational plans. In 1964, researchers ran a follow-up study on one-third of those seniors (10,317) and were able to contact 87% of that sample. In the next wave, in 1975, researchers interviewed 87% of the original 1964 sample by phone, including some people who had not responded in 1964. In fact, by 1975, a few members of the class of 1957 had died and the 87% reinterview rate was 93% of everyone who was still available then. In 1992, the researchers tracked the 9,741 survivors of the original 10,317 and interviewed 87% of them by phone for an hour.

It's not easy to track the participants in a longitudinal survey. Some die; some move; some simply disappear. But the effort to really track and interview the participants in a long-term project really pays off. (See Hauser [2005] for details on how the WLS research team has tracked the members of their original sample.) The WLS team began a new round of interviews in 2003 (the interviews were still going on in 2005 when I wrote this), and they are planning another wave for 2022, when the Wisconsin high school class of 1957 will be 83 years old (*ibid.*). The nonsensitive data from the WLS are available at <http://dpls.dacc.wisc.edu/wls/> and the sensitive data (about, for example, sexual preference, addiction, mental health, or criminal behavior) are available to qualified researchers at <http://www.ssc.wisc.edu/cdha/data/data.html> (Hauser 2005).

Perhaps the most famous panel study of all time is the Framingham Heart Study. In 1948, medical researchers began tracking 5,209 men and women between 30 and 62 years old from one small town—Framingham, Massachusetts. In 1971, as the original panel began to die off, another 5,124 panelists were added—this time, the original panelists' adult children and their spouses. Every 2 years, all the panelists go in for a complete medical check-up. This study has identified and nailed down the major risk factors for heart disease, which include, of course, behaviors (exercise, smoking) and inner states (attitudes, stress)—things that anthropologists, as well as epidemiologists, are interested in. Basic information about the Framingham study is available from the National Heart, Lung, and Blood Institute at <http://www.nhlbi.nih.gov/about/framingham/index.html> and on the study's website at <http://www.framingham.com/heart/>.

Like these long-term panel studies, longitudinal research by anthropologists are great treasures. Beginning in 1961, Robert Edgerton studied a sample of 48 mildly retarded people in California who had been released from a state mental institution (Edgerton 1967). This was full-blown participant observation: hanging out, following people around, doing in-depth interviews, taking field notes. Edgerton was able to interview 30 of the same people in 1975

(Edgerton and Bercovici 1976) and 15 members of the sample in 1982 (Edgerton et al. 1984). Edgerton last interviewed “Richard” in 1988, just before Richard died at age 68 (Edgerton and Ward 1991). As a result, we know more about how the mildly retarded get through life—how they make ends meet, how they deal (or don’t deal) with personal hygiene; how they get people to do things for them, like write letters; how people take advantage of them financially—than we could learn from any cross-sectional study.

Two of the best-known longitudinal studies in anthropology are the Tzintzuntzán project in Mexico and the Gwembe Tonga project in Zambia.

George Foster first went to Tzintzuntzán in 1945 to train some anthropology students from ENAH, Mexico’s National School of Anthropology and History (Foster 2002:254). Since then, either he or some other anthropologist has visited the community of 3,600 people almost every year. Foster’s students, Robert Van Kemper and Stanley Brandes, began going to Tzintzuntzán in 1967 and a third generation of students has already completed two doctoral dissertations there (Cahn 2002; Kemper and Royce 2002:192). Six comprehensive censuses have been taken in Tzintzuntzán from 1945 to 2000, and the files now include data on over 3,000 migrants who left Tzintzuntzán and live in the United States (Kemper 2002:303). (For more on this project and on data sources, see <http://www.santafe.edu/tarasco/>.)

The Gwembe Tonga Project began with visits in 1956 by Elizabeth Colson and Thayer Scudder. Some 57,000 Gwembe Tonga were being resettled to make way for the lake that would form in back of the Kariba dam on the Zambezi River, and Scudder and Colson were studying the effects of that resettlement. In 1962–1963, they realized the “long-term possibilities involved in a study of continuity and change among a people who, having been forcibly resettled in connection with a major dam, were soon to be incorporated with the independent nation of Zambia,” as the colonial period came to an end (Scudder and Colson 2002:200). Colson and Scudder continued their work and began recruiting colleagues into the project, including Lisa Cliggett, Sam Clark, and Rhonda Gillett-Netting, the three anthropologists who now manage the project (Cliggett 2002; Kemper and Royce 2002:192). Like the Tzintzuntzán project, the Gwembe project has incorporated indigenous members on the team.

Neither of these important projects started out as longitudinal studies. They just went on and on and on. The field notes and other data from these projects grow in importance every year, as more information is added to the corpus. All cross-sectional studies, including Master’s and Ph.D. projects, should be designed as if they were the start of a lifetime of research. You never know. (For more on panel studies, see Halaby 2004.)

ATTRITION

Panel studies often suffer from what's called **attrition** or the **respondent mortality** problem. This is where people drop out between successive waves of the panel survey. If this happens, and the results of successive waves are very different, you can't tell if that's because of (1) the special character of the drop out population, (2) real changes in the variables you're studying, or (3) both. For example, if dropouts tend to be male or poor, your results in successive waves will overrepresent the experiences of those who are female or affluent. If you run a longitudinal study, consult a statistician about how to test for the effects of attrition (see Rubin 1976; Fitzgerald et al. 1998; Thomas et al. 2001; Twisk and de Vente 2002. For an overview of the attrition problem, see Ahern and Le Brocq 2005.)

Respondent mortality is not always a problem. Roger Trent and I studied riders on Morgantown, West Virginia's "People Mover," an automated transport system that was meant to be a kind of horizontal elevator. You get on a little railway car (they carry only 8 seated and 12 standing passengers), push a button, and the car takes you to your stop—a block away or 8 miles across town. The system was brought on line a piece at a time between 1975 and 1980. Trent and I were tracking public support as the system went more places and became more useful (Trent and Bernard 1985). We established a panel of 216 potential users of the system when the system opened in 1975 and reinterviewed the members of that panel in 1976 and 1980 as more and more pieces of the system were added.

All 216 original members of the panel were available during the second wave and 189 were available for the third wave of the survey. Note, though, that people who were unavailable had moved out of Morgantown and were no longer potential users of the system. What counted in this case was maintaining a panel large enough to represent the attitudes of people in Morgantown about the People Mover system. The respondents who stayed in the panel still represented the people whose experiences we hoped to learn about.

Some Specialized Survey Techniques

Factorial Surveys

In a **factorial survey** (Rossi and Nock 1982; Rossi and Berk 1997), people are presented with **vignettes** that describe hypothetical social situations and are asked for their judgments about those situations. The General Social Survey is a face-to-face survey of about 1,600 adults that's done almost every year in the United States (for more on the GSS, go to <http://webapp.icpsr.umich.edu/GSS/>). Here's a vignette that was in the 1992 GSS:

This family has four children, the youngest is 6 months old, living with their mother. The mother is divorced. The mother has a college degree and is unemployed and not looking for work because she has no ready means of transportation. The father has remarried and is permanently disabled. The family is likely to face financial difficulties for a couple of years. Her parents cannot afford to help out financially. The family has \$1,000 in savings. All in all, the family's total income from other sources is \$100 per week.

What should this family's weekly income be? Include both the money already available from sources other than the government, and any public assistance support you think this family should get.

<i>Amount already received by this family</i>	<i>Average U.S. family income</i>
X	X
0 50 100 150 200 250 300 350 400 450 500 550 600	

SOURCE: Reprinted from *Social Science Research*, Vol. 22, No. 3, J. A. Will, "The Dimensions of Poverty: Public Perceptions of the Deserving Poor," p. 322, © 1993, with permission from Elsevier.

There are 10 variables in this vignette (number of children, marital status of the mother, how much savings the family has, the total income of the family, etc.), with 1,036,800 possible combinations. That seems just about right to me. The calculus for any individual's opinion about how much money to award the deserving poor on welfare is really that complicated. Now, each of the 1,600 respondents in the GSS saw seven vignettes each, so the survey captured:

$$(1,600 \text{ people}) (7 \text{ vignettes}) (10 \text{ variables}) = 112,000$$

combinations, or a sample of about 11% of all the factors that probably go into people's opinion on this issue.

The results of that survey were very interesting. Respondents awarded people who were looking for work a lot more than they awarded people who weren't looking for work. But mothers only got an extra \$6 per week for seeking work, while fathers got over \$12. And if mothers were unemployed because they wouldn't take minimum-wage jobs, they had their allotments reduced by \$20 per week, on average, compared to what people were willing to give mothers who were working full-time. (For comparison, it took about \$180 in 2005 to buy what \$100 bought in 1986.)

The factorial survey combines the validity of randomized experiments with the reliability of survey research and lets you measure subtle differences in opinion. (For more on factorial surveys, see Jasso 1998, Lauder et al. 2001, Brew and Cairns 2004, and Herzog 2004. For an interesting use of vignettes

to test the difference in men's and women's attitudes about rape in Turkey, see Gölge et al. 2003.)

Time Budgets and Diaries

Time budget surveys have been done all over the world to track how ordinary human beings spend most of their days (Szalai 1972). The idea is to learn about the sequence, duration, and frequency of behaviors and about the contexts in which behaviors take place. Some researchers ask respondents to keep diaries; others conduct "yesterday interviews," in which respondents are asked to go over the last 24 hours and talk about everything they did. Some researchers combine these methods, collecting diaries from respondents and then following up with a personal interview.

Perhaps the earliest time budget in anthropology was done by Audrey Richards between 1930 and 1934 during two and a half years of fieldwork among the Bemba of Zambia (it was Northern Rhodesia back then). Richards went across the country, spending from 3 to 6 weeks in a series of villages. She pitched her tent in the middle of each village so she could watch people's activities and in two of those villages, Kasaka and Kapamba, she kept daily calendars of all the adults, both men and women (Richards 1939:10–11). Richards asked several informants to estimate how long they took to accomplish various tasks—planting a garden, chasing locusts, cutting trees. Then she averaged what people told her and got a rough estimate for each task. Then she applied those figures to her observations of people in the two villages where she kept records of the daily activities (*ibid.*:395 and appendix E).

In 1992, Elizabeth Harrison recruited 16 farmers in two villages of Luapula Province of Zambia to keep records for 4 months of their daily activities. The diaries reveal that the technology for producing cassava meal hadn't changed since Richards's day (Harrison 2000:59). They also showed how long seemingly ordinary things can take. Here is Abraham Kasongo, one of Harrison's informants, describing his trip to Kalaba, the capital of the province, to get millet so his mother could brew beer. Kalaba is 8 kilometers away and Kasongo is going by bicycle:

22nd July 1992

Morning I go watering the seeds after watering I came back and wash my body and go to my father's house to get the biscley and start the journey to Kalaba to get the millet. I found the one who has been given the money is not around I start waiting for him around 14 hrs he came and give me millet I go where the people in the village where drinking the coll me and I join them around 15 hrs I start coming back I found my wife is not around I go to my father's house and put

millet then I show my father the fish for sale and the piace is K200.00 he take the fish and I start coming back straight to the house. I found my wife priparing fire and start cooking Nshima [the staple food, made from cassava. HRB] with dry vegetables we eat and I go to see Eliza we tolked antill I cam back to slip because I was tired I just go straight to slip. (Harrison 2000:62)

Discursive diaries, in other words, are like any other qualitative data: They make process clear and bring out subtleties in behavioral complexes that time budgets can obscure. And, just as with any other survey method, getting both the qualitative and quantitative is better than one kind of data alone.

Susan Shaw's study of family activities in Canada is typical of the use of time budgets in modern societies (1992). Shaw studied 46 middle- and working-class couples who had children living at home. All the fathers were employed full-time and among the mothers, 12 were employed full time, nine were employed part-time, and 25 were full-time homemakers. Both parents kept time diaries for 1 day during the week and for 1 day on a weekend. Then, Shaw interviewed the parents separately, for 1 to 2 hours in their homes. For each activity that they had mentioned, parents were asked if they considered the activity to be work or leisure, and why.

Shaw calculated the amount of time that each parent reported spending with their children—playing with them, reading to them, and so on. The rather dramatic results are in table 10.1. For these Canadian families, at least, the more

TABLE 10.1
Average Amount of Time Fathers and Mothers Report Spending with Children,
by the Mother's Employment Status

<i>Mother's employment status</i>	<i>N</i>	<i>Time with children per day (in minutes)</i>	
		<i>Mothers</i>	<i>Fathers</i>
Employed full-time	12	97	71
Employed part-time	9	144	52
Full-time homemaker	25	241	23
Total	46		

SOURCE: "Dereifying Family Leisure: An Examination of Women's and Men's Everyday Experiences and Perceptions of Family Time" by S. Shaw, 1992, *Leisure Sciences*, p. 279. Reproduced by permission of Taylor and Francis.

that women work outside the home, the more time fathers spend with their children.

Diaries and time-budget interviews, particularly with the aid of checklists, appear to be more accurate than 24-hour recall of activities. But no matter

what you call them, time budgets and diaries are still methods for collecting self-reports of behavior. They may be *less* inaccurate than simply asking people to tell you what they did over the past day or week, but they are not perfect. A lot of work remains to be done on testing the accuracy of activity diaries against data from direct observation. In chapter 15, we'll look at methods for direct observation and measurement of behavior.

Randomized Response

Randomized response is a technique for estimating the amount of some socially negative behavior in a population—things like shoplifting, extramarital sex, child abuse, being hospitalized for emotional problems, and so on. The technique was introduced by Warner in 1965 and is particularly well described by Williams (1978:73). It is a simple, fun, and interesting tool. Here's how it works.

First, you formulate two questions, *A* and *B*, that can be answered “yes” or “no.” One question, *A*, is the question of interest (say, “Have you ever shoplifted?”) The possible answers to this question (either “yes” or “no”) do not have known probabilities of occurring. That is what you want to find out.

The other question, *B*, must be innocuous and the possible answers (again “yes” or “no”) must have known probabilities of occurring. For example, if you ask someone to toss a fair coin and ask, “Did you toss a heads?” then the probability that they answer “yes” or “no” is 50%. If the chances of being born in any given month were equal, then you could ask respondents: “Were you born between April 1st and June 1st?” and the probability of getting a “yes” would be 25%. Unfortunately, births are seasonal, so the coin-toss question is preferable.

Let's assume you use the coin toss for question *B*. You ask someone to toss the coin and to note the result *without letting you see it*. Next, have them pick a card, from a deck of 10 cards, where each card is marked with a single integer from 1 to 10. The respondent *does not tell you what number he or she picked*, either. The secrecy associated with this procedure makes people feel secure about answering question *A* (the sensitive question) truthfully.

Next, hand the respondent a card with the two questions, marked *A* and *B*, written out. Tell them that if they picked a number between one and four from the deck of 10 cards, they should answer question *A*. If they picked a number between five and 10, they should answer question *B*.

That's all there is to it. You now have the following: (1) Each respondent knows they answered “yes” or “no” and which question they answered; and (2) You know *only* that a respondent said “yes” or “no” but not which question, *A* or *B*, was being answered.

If you run through this process with a sufficiently large, representative sample of a population, and if people cooperate and answer all questions truthfully, then you can calculate the percentage of the population that answered “yes” to question *A*. Here’s the formula:

$$P_{A \text{ or } B} = [(P_{A+} \times P_A) + (P_{B+} \times P_B)] \quad \text{Formula 10.2}$$

The percentage of people who answer “yes” to *either A or B* = (the percentage of people who answer “yes” to question *A*) times (the percentage of times that question *A* is asked) plus (the percentage of people who answered “yes” to question *B*) times (the percentage of times question *B* is asked).

The only unknown in this equation is the percentage of people who answered “yes” to question *A*, the sensitive question. We know, from our data, the percentages of “yes” answers to *either* question. Suppose that 33% of all respondents said “yes” to *something*. Since respondents answered question *A* only if they chose a number from 1 to 4, then *A* was answered 40% of the time and *B* was answered 60% of the time. Whenever *B* was answered, there was a 50% chance of it being answered “yes” because that’s the chance of getting a heads on the toss of a fair coin. The problem now reads:

$$\begin{aligned} .33 &= A(.40) + .50(.60) \text{ or} \\ .33 &= .40A + .30 \end{aligned}$$

which means that $A = .075$. That is, given the parameters specified in this experiment, if 33% of the sample says “yes” to either question, then about 8% of the sample answered “yes” to question *A*.

There are two problems associated with this technique. First, no matter what you say or do, some people will not believe that you can’t identify them and will therefore not tell the truth. Bradburn, Sudman et al. (1979) report that 35% of known offenders would not admit to having been convicted of drunken driving in a randomized response survey. Second, like all survey techniques, randomized response depends on large, representative samples. Since the technique is time consuming to administer, this makes getting large, representative samples difficult.

Still, the evidence is mounting that for some sensitive questions—Did you smoke dope in the last week? Have you ever bought a term paper? Have you stolen anything from your employer?—when you want the truth, the randomized response method is worth the effort (see Scheers and Dayton [1987], Nordlund et al. [1994], and Clark and Desharnais [1998] for some examples).

Every time I read in the newspaper that self-reported drug use among adolescents has dropped by such-and-such and amount since whenever the last

self-report survey was done, I think about how easy it is for those data to be utter nonsense. And I wonder why the randomized response technique isn't more widely used.

Dietary Recall

Studies of diet and human nutrition mostly rely on informants to recall what they've eaten over the last 24 hours or what they usually eat for various meals. These **24-hour recall** tests, or **dietary recall** interviews, are a specialized kind of structured interview. The problem is, they often produce dreadfully inaccurate results.

C. J. Smith et al. (1996) compared the responses of 575 Pima and Papago Indians (in Arizona) to a 24-hour recall instrument about food intake with responses to a very detailed survey called the Quantitative Food Frequency questionnaire. In the QFF, interviewers probe for a list of regularly consumed foods in a community. Smith et al. also assessed the energy expenditure of 21 people in the research group using the doubly labeled water technique. The DLW technique involves giving people special water to drink—water with isotopes that can be tracked in blood and urine samples—and then testing, over time, their actual intake of nutrients.

The correlation, across the 21 participants, between the energy intake measured by the DLW technique and the energy intake estimated by the informants' responses to the QFF, was 0.48. This correlation is statistically significant, but it means that just 23% (0.48^2) of the variation in actual energy intake across the 21 people was accounted for by their responses to a very detailed interview about their food consumption. And the correlation of actual energy intake with estimates from the 24-hour recall data was much worse.

Johnson et al. (1996) also found no useful relation between individual 24-hour recall measurements of energy intake among children in Vermont and measurements of those same children by the DLW technique. But, in the all-is-not-lost department, Johnson et al. found that averaging the data for energy intake across *three* 24-hour recalls in 14 days (on day 1, day 8, and day 14) produced results that were very similar to those produced by the DLW technique. So, people hover around giving accurate answers to a question about calorie intake and if you get at least three answers, for three time windows, and take the average, you may get a useful result. (For more methods in nutritional anthropology, see Quandt and Ritenbaugh 1986 and Pelto et al. 1989. For more on measuring food intake, see Chapman et al. 1997, Schoenberg 1997, 2000, Melnik et al. 1998, and Graham 2003.)

Mixed Methods

Finally, this: With all the great techniques out there for collecting systematic data, there is nothing to stop you from using several methods, even wildly different methods like narratives, questionnaires, and randomized response, in the same study. By now, you know that there is no need to choose between qualitative and quantitative data. Whether you are doing exploratory or confirmatory research, a sensible mix of methods—methods that match the needs of the research—is what you're after.

Furthermore, there is no formula for how to mix methods. You'll use ethnography to develop good questions for a questionnaire, but you'll also use ethnography to interpret and flesh out the results from questionnaires. Ethnography can tell you what parameters you want to estimate, but you need survey data to actually estimate parameters. Ethnography brings to light the features of a culture, but you need systematically collected data (surveys that produce either words or numbers) in order to test hypotheses about how those features work. Researchers who are comfortable with both will routinely move back and forth, without giving it a moment's thought. Ethnography tells you that patrilineal cross-cousin marriage is preferred, but it takes a survey to find out how often the rule is ignored. And then it takes more ethnography to find out how people rationalize ignoring the cultural preference.

Today, mixed methods is becoming the norm rather than something interesting to talk about. Not a moment too soon, either. (On mixing qualitative and quantitative methods, see Pearce 2002, Tashakkori and Teddlie 2003, and Mertens 2005.)



Structured Interviewing II: Cultural Domain Analysis

Cultural Domain Analysis

Cultural domain analysis is the study of how people in a group think about lists of things that somehow go together. These can be lists of physical, observable things—plants, colors, animals, symptoms of illness—or conceptual things—occupations, roles, emotions. The goal is to understand how people in different cultures (or subcultures) interpret the content of domains differently (Borgatti 1993/1994).

The spectrum of colors, for example, has a single physical reality that you can see on a machine. Some peoples across the world, however—Xhosa, Navajo, *Ñähñu*—identify the colors across the physical spectrum of green and blue with a single gloss. In *Ñähñu*, for example, the word is *nk'ami* and in Navajo it's *dootl'izh*. Linguists who study this phenomenon call this color “grue” (see, for example, Branstetter 1977, Kim 1985, and Davies et al. 1994).

This does *not* mean that people who have a word for grue fail to *see* the difference between things that are the color of grass and things that are the color of a clear sky. They just *label* chunks of the physical spectrum of colors differently than we do and use adjectival modifiers of grue to express color differences within the blue-green spectrum. In Navajo, turquoise is *yáago dootl'izh*, or “sky grue,” and green is *tádlidgo dootl'izh*, or “water skum grue” (Oswald Werner, personal communication). If this seems exotic to you, get a chart of, say, 100 lipstick colors or house paint colors and ask people at your university to name the colors. On average, women will probably recog-

nize (and name) more colors than men will; and art majors of both sexes will name more colors than, say, engineering majors will.

This concern for understanding cultural differences in how people cut the natural world goes a long way back in anthropology. Lewis Henry Morgan (1997 [1870]) studied systems of kinship nomenclature. His work made clear that if someone says, “This is my sister,” you can’t assume that they have the same mother and father. Lots of different people can be called “sister,” depending on the kinship system.

In his work with the Murray Islanders (in the Torres Straits between Australia and Papua New Guinea) and then later with the Todas of southern India, W.H.R. Rivers developed the genealogical method—those ego-centered graphs for organizing kinship data that we take for granted today—as a way to elicit accurately and systematically the inventory of kin terms in a language (Rivers 1906, 1910, 1968 [1914]; and see Rivers’s work in Volume VI of Haddon 1901–1935).

Anthropologists also noticed very early that, although kinship systems *could* be unique to each culture—which would mean that each system required a separate set of rules—they simply weren’t. Alfred Kroeber showed in 1909 that just eight features were needed to distinguish kinship terms in any system: (1) whether the speaker and the kin referred to were of the same or different generations; (2) the relative age of people who are of the same generation—older or younger brother, for example; (3) whether the person referred to is a collateral or a lineal relative; (4) whether the person referred to is an affinal or consanguineal relative; (5) whether the relative is male or female; (6) whether the speaker is male or female; (7) whether the person who links the speaker and the relative is male or female; and (8) whether the person who links the speaker and the relative is alive or dead.

Now, if you first choose whether to use or not use any of those eight features and then choose among the two alternatives to each feature, you can concoct $3^8 = 6,561$ kinds of kinship systems. But, while there are some rare exceptions (the bilineal Yakö of Nigeria, the ambilineal Gilbert Islanders), most of the world’s kinship systems are of one those familiar types you studied in Anthropology 101—the Hawaiian, Sudanese, Omaha, Eskimo, Crow, and Iroquois types. Early anthropologists found it pretty interesting that the world’s real kinship systems comprised just a tiny set of the possibilities. Ever since the work of Morgan and Kroeber and Rivers, a small, hardy band of anthropologists has tried to determine if these systems are associated with particular political, economic, or environmental conditions (Leach 1945; Alexander 1976; Lehman 1992; Houseman and White 1998; Kronenfeld 2004).

This interest in classifying kinship systems led to methods for discovering sets of terms in other domains, like kinds of foods, things to do on the week-

end, kinds of crime, bad names for ethnic groups, dirty words, names for illnesses, etc. Note that none of these is about people's preferences. Asking people to tell you "which animals do you think make good pets" is very different from asking them to "list animals that people here keep as pets" (Borgatti 1999:117).

We usually ask people about their preferences because we want to predict those preferences. If we ask people which of two political candidates they favor in an election, for example, we might also ask them about their income, their ethnicity, their age, and so on. Then we look for packages of variables about the people that predict their preference for a candidate. Or we might do the same thing to predict why people prefer certain brands of cars, or why they have this or that position on controversial issues. In cultural domain analysis, however, we're interested in the items that comprise the domain—the illnesses, the edible plants, the jobs that women and men do, etc.

In other words, we're interested in things external to the people we interview and *how those things are related to each other* in people's minds (Spradley 1979; Borgatti 1999). For example, things can be *kinds of* other things: an orange is a kind of fruit, and a Valencia is a kind of orange. Cultural domain analysis involves, among other things, the building of folk taxonomies from data that informants supply about what goes with what. I'll show you how to build folk taxonomies in chapter 18 when we get to methods for analyzing qualitative data. Here, I want to focus on methods for collecting lists and similarities among the items in a list—that is, the contents of a domain and people's ideas about what goes with what. These methods include *free lists*, *sentence frames*, *triad tests*, *pile sorts*, *paired comparisons*, *rankings*, and *rating scales*. The last of these, rating scales, is a major field of measurement all by itself and so it gets its own chapter, right after this one.

Two things make structured interviewing methods very productive. First, they are fun to administer and fun for people to do. Second, *Anthropac* software (Borgatti 1992a, 1992b) makes it easy to collect and analyze data using these techniques.

Free Listing

Free listing is a deceptively simple, but powerful technique. In free listing, you ask informants to "list all the X you know about" or "what kinds of X are there?" where X might be movie stars, brands of computers, kinds of motor vehicles, etc.

The object is to get informants to list as many items as they can in a domain, so you need to probe and not just settle for whatever people say. Brewer et al.

(2002) found that you can increase recall with four kinds of probes: (1) redundant questioning, (2) nonspecific prompting, (3) prompting with alphabetic cues, and (4) prompting with semantic cues. Here's the redundant question that Brewer and his colleagues asked a group of IV-drug users:

Think of all the different kinds of drugs or substances people use to get high, feel good, or think and feel differently. These drugs are sometimes called recreational drugs or street drugs. Tell me the names of all the kinds of these drugs you can remember. Please keep trying to recall if you think there are more kinds of drugs you might be able to remember. (ibid.:347)

Notice how the question is repeated in different words and with a few cues built in. In nonspecific prompting you ask people "What other kinds of X are there?" after they've responded to your original question. You keep asking this question until people say they can't think of any more Xs.

With alphabetic cues, you go through the alphabet and ask informants "what kinds of X are there that begin with the letter A?" And in semantic cues, you take the first item in an informant's initial list and ask: "Think of all the kinds of X that are like Y," where Y is that first item on the list. "Try to remember other types of X like Y and tell me any new ones that you haven't already said." You do this for the second item, the third and so on. This technique increases recall of items by over 40% (Brewer et al. 2002:112).

You'd be surprised at how much you can learn from a humble set of free lists. Henley (1969) asked 21 adult Americans (students at Johns Hopkins University) to name as many animals as they could in 10 minutes. She found an enormous variety of expertise when it comes to naming animals. In just this small group of informants (which didn't even represent the population of Johns Hopkins University, much less that of Baltimore or the United States), the lists ranged in length from 21 to 110, with a median of 55.

In fact, those 21 people named 423 different animals, and 175 were mentioned just once. The most popular animals for this group of informants were: dog, lion, cat, horse, and tiger, all of which were named by more than 90% of informants. Only 29 animals were listed by more than half the informants, but 90% of those were mammals. By contrast, among the 175 animals named only once, just 27% were mammals.

But there's more. Previous research had shown that the 12 most commonly talked about animals in American speech are: bear, cat, cow, deer, dog, goat, horse, lion, mouse, pig, rabbit, and sheep. There are $n(n - 1)/2$, or 66 possible unique pairs of 12 animals (dog-cat, dog-deer, horse-lion, mouse-pig, etc.). Henley examined each informant's list of animals, and found the difference in the order of listing for each of the 66 pairs.

That is, if an informant mentioned goats 12th on her list, and bears 32nd, then the distance between goats and bears, for that informant, was $32 - 12 = 20$. Henley standardized these distances (that is, she divided each distance by the length of an informant's list and multiplied by 100) and calculated the average distance, over all the informants, for each of the 66 pairs of animals.

The lowest mean distance was between sheep and goats (1.8). If you named sheep, then the next thing you named was probably goats; and if you named goats, then next thing you named was probably sheep. Most speakers of English (and all other major Western languages, for that matter) have heard the expression: "That'll separate the sheep from the goats."

This part of Western culture was originally a metaphor for distinguishing the righteous from the wicked and then became a metaphor for separating the strong from the weak. The first meaning was mentioned in the Old Testament (Ezekiel 34:17), and then again around 600 years later in the New Testament (Matthew 25:31–33). Nowadays, you might hear someone say "Boy, *that'll* separate the sheep from the goats" on their way out of a calculus exam.

Henley's respondents were neither shepherds nor students of Western scriptural lore, but they all knew that sheep and goats somehow "go together." Free lists tell you *what goes with what*, but you need to dig in order to understand *why*. Cats and dogs were only 2 units apart in Henley's free lists—no surprise there, right?—while cats and deer were 56 units apart. Deer, in fact, are related to all the other animals on the list by at least 40 units of distance, except for rabbits, which are only 20 units away from deer.

Robert Trotter (1981) reports on 378 Mexican Americans who were asked to name the *remedios caseros*, or home remedies, they knew, and what illnesses each remedy was for. Informants listed a total of 510 remedies for treating 198 illnesses. However, the 25 most frequently mentioned remedies—about 5% of the 510—made up about 41% of all the cases; and the 70 most frequently mentioned illnesses—about 14%—made up 84% of the cases.

Trotter's free-list data reveal a lot about Mexican American perceptions of illness and home cures. He was able to count which ailments were reported more frequently by men and which by women; which ailments were reported more frequently by older people and which by younger people; which by those born in Mexico and which by those born in the United States; and so on.

Free listing is often a prelude to cluster analysis and multidimensional scaling, which we'll get to in chapter 21. But consider what John Gatewood (1983a) learned from *just* a set of free lists. He asked 40 adult Pennsylvanians to name all the trees they could think of. Then he asked them to check the trees on their list that they thought they could recognize in the wild. Thirty-seven of them listed "oak," 34 listed "pine," 33 listed "maple," and 31 listed "birch." I suspect that the list of trees and what people say they could recog-

nize would look rather different in, say Wyoming or Mississippi. We could test that.

Thirty-one of the 34 who listed “pine” said they could recognize a pine. Twenty-seven people listed “orange,” but only four people said they could recognize an orange tree (without oranges hanging all over it, of course). On average, the Pennsylvanians in Gatewood’s sample said they could recognize half of the trees they listed. Gatewood calls this the **loose talk phenomenon**. He thinks that many Americans can name a lot more things than they can recognize in nature.

Does this loose talk phenomenon vary by gender? Suppose, Gatewood says, we ask Americans from a variety of subcultures and occupations to list other things besides trees. Would the 50% recognition rate hold?

Gatewood and a group of students at Lehigh University asked 54 university students, half women and half men, to list all the musical instruments, fabrics, hand tools, and trees they could think of. Then the informants were asked to check off the items in each of their lists that they thought they would recognize in a natural setting.

Gatewood chose musical instruments, with the idea that there would be no gender difference in the number of items listed or recognized; he thought that women might name more kinds of fabrics than would men and that men would name more kinds of hand tools than would women. He chose the domain of trees to see if his earlier findings would replicate. All the hypotheses were supported (Gatewood 1984).

A. Kimball Romney and Roy D’Andrade asked 105 American high school students to “list all the names for kinds of relatives and family members you can think of in English” (1964:155). They were able to do a large number of analyses on these data. For example, they studied the order and frequency of recall of certain terms, and the productiveness of modifiers, such as “step-,” “half-,” “-in-law,” “grand-,” “great-,” and so on. They assumed that the nearer to the beginning of a list a kin term occurs, the more salient it is for that particular informant. By taking the average position in all the lists for each kin term, they were able to derive a rank order list of kin terms, according to the variable’s saliency.

They also assumed that more salient terms occur more frequently. So, for example, “mother” occurs in 93% of all lists and is the first term mentioned on most lists. At the other end of the spectrum is “grandson,” which was only mentioned by 17% of the 105 informants, and was, typically, the 15th, or last term to be listed. They found that the terms “son” and “daughter” occur on only about 30% of the lists. But remember, these informants were all high school students, all of whom *were* sons and daughters, but none of whom *had* sons or daughters. It would be interesting to repeat Romney and D’Andrade’s

experiment on many different American populations. We could then test the saliency of English kin terms on the many subpopulations.

Finally, free listing can be used to find out where to concentrate effort in applied research, and especially in rapid assessment. Researchers interested in high-risk sexual behavior, for example, use the free-list technique to understand domains like “ways to have sex” (Schensul et al. 1994) and “reasons to have sex” (Flores et al. 1998).

Monárrez-Espino et al. (2004) worked on a food aid program for at-risk Tarahumara infants in Mexico. A government agency had developed a basket of nutritional foods for distribution to Tarahumara mothers, but many of the foods (like canned sardines) were culturally unacceptable. Free listing of foods helped set things right.

In a project on which I consulted, interviewers asked people on the North Carolina coast how they viewed the possibility of offshore oil drilling. One of the questions was: “What are the things that make life good around here?” This question cropped up after some informal interviews in seven small, seaside towns. People kept saying “What a nice little town this is” and “What a shame it would be if things changed around here.” Informants had no difficulty with the question, and after just 20 interviews, the researchers had a list of over 50 “things that make life good around here.” The researchers chose the 20 items mentioned by at least 12 informants and explored the meaning of those items further (ICMR et al. 1993).

The humble free list has many uses. Use it a lot.

The True-False/Yes-No and Sentence Frame Techniques

Another common technique in cultural domain analysis is called the **sentence frame** or **frame elicitation** method. Linda Garro (1986) used the frame elicitation method to compare the knowledge of curers and noncurers in Pichátaro, Mexico. She used a list of 18 illness terms and 22 causes, based on prior research in Pichátaro (Young 1978). The frames were questions, like “can _____ come from _____?” Garro substituted names of illnesses in the first blank, and things like “anger,” “cold,” “overeating,” and so on in the second blank. (*Anthropac* has a routine for building questionnaires of this type.) This produced an 18 × 22 yes-no matrix for each of the informants. The matrices could then be added together and submitted to analysis by multi-dimensional scaling (see chapter 21).

James Boster and Jeffrey Johnson (1989) used the frame-substitution method in their study of how recreational fishermen in the United States categorize ocean fish. They asked 120 fishermen to consider 62 belief frames, scan

down a list of 43 fish (tarpon, silver perch, Spanish mackerel, etc.), and pick out the fish that fit each frame. Here are a few of the belief frames:

The meat from _____ is oily tasting.
 It is hard to clean _____.
 I prefer to catch _____.

That's $43 \times 62 = 2,666$ judgments by each of 120 informants, but informants were usually able to do the task in about half an hour (Johnson, personal communication). The 62 frames, by the way, came straight out of ethnographic interviews where informants were asked to list fish and to talk about the characteristics of those fish.

Gillian Sankoff (1971) studied land tenure and kinship among the Buang, a mountain people of northeastern New Guinea. The most important unit of social organization among the Buang is the *dgwa*, a kind of descent group, like a clan. Sankoff wanted to figure out the very complicated system by which men in the village of Mambump identified with various *dgwa* and with various named garden plots.

The Buang system was apparently too complex for bureaucrats to fathom, so, to save administrators a lot of trouble, the men of Mambump had years earlier devised a simplified system that they presented to outsiders. Instead of claiming that they had ties with one or more of five different *dgwa*, they each decided which of the two largest *dgwa* they would belong to, and that was as much as the New Guinea administration knew.

To unravel the complex system of land tenure and descent, Sankoff made a list of all 47 men in the village and all 140 yam plots that they had used over the recent past. Sankoff asked each man to go through the list of men and identify which *dgwa* each man belonged to. If a man belonged to more than one, then Sankoff got that information, too. She also asked her informants to identify which *dgwa* each of the 140 garden plots belonged to.

As you might imagine, there was considerable variability in the data. Only a few men were uniformly placed into one of the five *dgwa* by their peers. But by analyzing the matrices of *dgwa* membership and land use, Sankoff was able to determine the core members and peripheral members of the various *dgwa*.

She was also able to ask important questions about intracultural variability. She looked at the variation in cognitive models among the Buang for how land use and membership in descent groups were related. Sankoff's analysis was an important milestone in our understanding of the measurable differences between individual culture vs. shared culture. It supported Goodenough's

notion (1965) that cognitive models are based on shared assumptions, but that ultimately they are best construed as properties of individuals.

Techniques like true-false and yes-no tests that generate nominal data are easy to construct, especially with *Anthropac*, and can be administered to a large number of informants. Frame elicitation in general, however, can be boring, both to the informant and to the researcher alike. Imagine, for example, a list of 25 animals (mice, dogs, antelopes . . .), and 25 attributes (ferocious, edible, nocturnal . . .).

The structured interview that results from such a test involves a total of 625 (25×25) questions to which an informant must respond—questions like “Is an antelope edible?” “Is a dog nocturnal?” “Is a mouse ferocious?” People can get pretty exasperated with this kind of foolishness, so be careful to choose domains, items, and attributes that make sense to people when you do frame elicitations and true-false tests.

Triad Tests

In a **triad test**, you show people three things and tell them to “Choose the one that doesn’t fit” or “Choose the two that seem to go together best,” or “Choose the two that are the same.” The “things” can be photographs, dried plants, or 3" \times 5" cards with names of people on them. (Respondents often ask “What do you mean by things being ‘the same’ or ‘fitting together’?” Tell them that you are interested in what *they* think that means.) By doing this for all triples from a list of things or concepts, you can explore differences in cognition among individuals, and among cultures and subcultures.

Suppose you ask a group of Americans to “choose the item that is least like the other two” in each of the following triads:

DOLPHIN	MOOSE	WHALE
SHARK	DOLPHIN	MOOSE

All three items in the first triad are mammals, but two of them are sea mammals. Some native speakers of English will choose “dolphin” as the odd item because “whales and moose are both big mammals and the dolphin is smaller.” In my experience, though, most people will choose “moose” as the most different because “whales and dolphins are both sea animals.” In the second triad, many of the same people who chose “moose” in the first triad will choose “shark” because moose and dolphins are both mammals and sharks are not.

But some people who chose “moose” in triad 1 will choose “moose” again

because sharks and dolphins are sea creatures, while moose are not. Giving people a judiciously chosen set of triad stimuli can help you understand inter-individual similarities and differences in how people think about the items in a cultural domain.

The triads test was developed in psychology (see Kelly 1955; Torgerson 1958) and has long been used in studies of cognition. Romney and D'Andrade (1964) presented people with triads of American kinship terms and asked them to choose the term that was most *dissimilar* in each triad. For example, when they presented informants with the triad “father, son, nephew,” 67% selected “nephew” as the most different of the three items. Twenty-two percent chose “father” and only 2% chose “son.” Romney and D'Andrade asked people to explain *why* they'd selected each item on a triad. For the triad “grandson, brother, father,” for example, one informant said that a “grandson is most different because he is moved down further” (p. 161). There's a lot of cultural wisdom in that statement.

By studying which pairs of kinship terms their informants chose most often as being as similar, Romney and D'Andrade were able to isolate some of the salient components of the American kinship system (components such as male vs. female, ascending vs. descending generation, etc.). They were able to do this, at least, for the group of informants they used. Repeating their tests on other populations of Americans, or on the same population over time, would yield interesting comparisons.

Lieberman and Dressler (1977) used triad tests to examine intracultural variation in ethnomedical beliefs on the Caribbean island of St. Lucia. They wanted to know if cognition of disease terms varied with bilingual proficiency. They used 52 bilingual English-Patois speakers, and 10 monolingual Patois speakers. From ethnographic interviewing and cross-checking against various informants, they isolated nine disease terms that were important to St. Lucians.

Here's the formula for finding the number of triads in a list of n items:

$$\text{The number of triads in } n \text{ items} = \frac{n(n-1)(n-2)}{6} \quad \textbf{Formula 11.1}$$

In this case, $n = 9$, so there are 84 possible triads.

Lieberman and Dressler gave each of the 52 bilingual informants two triad tests, a week apart: one in Patois and one in English. (Naturally, they randomized the order of the items within each triad, and also randomized the order of presentation of the triads to informants.) They also measured how bilingual their informants were, using a standard test. The 10 monolingual Patois informants were simply given the Patois triad test.

The researchers counted the number of times that each possible pair of

terms was chosen as most alike among the 84 triads. (There are $n(n - 1)/2$ pairs = $9(8)/2 = 36$ pairs.) They divided the total by seven (the maximum number of times that any pair appears in the 84 triads). This produced a similarity coefficient, varying between 0.0 and 1.0, for each possible pair of disease terms. The larger the coefficient for a pair of terms, the closer in meaning are the two terms. They were then able to analyze these data among English-dominant, Patois-dominant, and monolingual Patois speakers.

It turned out that when Patois-dominant and English-dominant informants took the triad test in English, their cognitive models of similarities among diseases was similar. When Patois-dominant speakers took the Patois-language triad test, however, their cognitive model was similar to that of monolingual Patois informants.

This is a very interesting finding. It means that Patois-dominant bilinguals manage to hold on to two distinct psychological models about diseases and that they switch back and forth between them, depending on what language they are speaking. By contrast, the English-dominant group displayed a similar cognitive model of disease terms, irrespective of the language in which they are tested.

The Balanced Incomplete Block Design for Triad Tests

Typically, the terms that go into a triad test are generated by a free list, and typically the list is much too long for a triad test. As you can see from Formula 11.1, with just nine terms, there are 84 stimuli in a triad test containing nine items. But with 15 items, just 6 more, the number of decisions an informant has to make jumps to 455. At 20 items, it's a mind-numbing 1,140.

Free lists of illnesses, ways to prevent pregnancy, advantages of breast-feeding, places to go on vacation, and so on easily produce 60 items or more. Even a selected, abbreviated list may be 20 items.

This led Michael Burton and Sara Nerlove (1976) to develop the **balanced incomplete block design**, or BIB, for the triad test. BIBs take advantage of the fact that there is a lot of redundancy in a triad test. Suppose you have just four items, 1, 2, 3, 4 and you ask informants to tell you something about *pairs* of these items (e.g., if the items were vegetables, you might ask "Which of these two is less expensive?" or "Which of these two is more nutritious?" or "Which of these two is easier to cook?") There are exactly six pairs of four items (1-2, 1-3, 1-4, 2-3, 2-4, 3-4), and the informant sees each pair just once.

But suppose that instead of pairs you show the informant triads and ask which two out of each triple are most similar. There are just four triads in four items (1-2-3, 1-2-4, 2-3-4, 1-3-4), but each item appears $(n - 1)(n - 2)/2$

times, and each pair appears $n - 2$ times. For four items, there are $n(n - 1)/2 = 6$ pairs; each pair appears twice in four triads, and each item on the list appears three times.

It is all this redundancy that reduces the number of triads needed in a triads test. In a complete set of 84 triads for nine items, each pair of items appears $n - 2$, or seven times. If you have each pair appear just once (called a **lambda 1 design**), instead of seven times, then, instead of 84 triads, only 12 are needed. If you have each pair appear twice (a lambda 2 design), then 24 triads are needed. For analysis, a lambda 2 design is much better than a lambda 1. Table 11.1 shows the lambda 2 design for 9 items and 10 items.

TABLE 11.1
Balanced Incomplete Block Designs for Triad Tests Involving 9 and 10 Items

<i>For 9 items, 24 triads are needed, as follows:</i>		<i>For 10 items, 30 triads are needed, as follows:</i>	
Items		Items	
1, 5, 9	1, 2, 3	1, 2, 3	6, 8, 9
2, 3, 8	4, 5, 6	2, 5, 8	7, 10, 3
4, 6, 7	7, 8, 9	3, 7, 4	8, 1, 10
2, 6, 9	1, 4, 7	4, 1, 6	9, 5, 2
1, 3, 4	2, 5, 9	5, 8, 7	10, 6, 7
5, 7, 8	3, 6, 8	6, 4, 9	1, 3, 5
3, 7, 9	1, 6, 9	7, 9, 1	2, 7, 6
2, 4, 5	2, 4, 8	8, 10, 2	3, 8, 9
1, 6, 8	3, 5, 7	9, 3, 10	4, 2, 10
4, 8, 9	1, 5, 8	10, 6, 5	5, 6, 3
3, 5, 6	2, 6, 8	1, 2, 4	6, 1, 8
1, 2, 7	3, 4, 9	2, 3, 6	7, 9, 2
		2, 4, 8	8, 4, 7
		4, 9, 5	9, 10, 1
		5, 7, 1	10, 5, 4

SOURCE: Reprinted from *Social Science Research*, Vol. 5, M. L. Burton and S. B. Nerlove, "Balanced Design for Triad Tests," p. 5, © 1976, with permission from Elsevier.

For 10 items, a lambda 2 design requires 30 triads; for 13 items, it requires 52 triads; for 15 items, 70 triads; for 19 items, 114 triads; and for 25 items, 200 triads. Unfortunately, there is no easy formula for choosing *which* triads in a large set to select for a BIB. Fortunately, Burton and Nerlove (1976) worked out various lambda BIB designs for up to 25 items and Borgatti has incorporated BIB designs into *Anthropac* (1992a, 1992b). You simply tell *Anthropac* the list of items you have, select a design, and tell it the number of informants you want to interview. *Anthropac* then prints out a randomized

triad test, one for each informant. (Randomizing the order in which the triads appear to informants eliminates “order-effects”—possible biases that come from responding to a list of stimuli in a particular order.)

Boster et al. (1987) used a triad test in their study of the social network of an office. There were 16 employees, so there were 16 “items” in the cultural domain (“the list of all the people who work here” is a perfectly good domain). A lambda 2 test with 16 items has 80 distinct triads. Informants were asked to “judge which of three actors was the most different from the other two.”

Triad tests are easy to create with *Anthropac*, easy to administer, and easy to score, but they can only be used when you have relatively few items in a cultural domain. In literate societies, most informants can respond to 200 triads in less than half an hour, but it can be a really boring exercise, and boring your informants is a really bad idea. I find that informants can easily handle lambda 2 triad tests with up to 15 items and 70 triads. But I also find that people generally prefer—even like—to do pile sorts.

Free Pile Sorts

In 1966, John Brim put the names of 58 American English role terms (mother, gangster, stockbroker, etc.) on slips of paper. He asked 108 high school students in San Mateo, California, to spread the slips out on their desks and to “put the terms together which you feel belong together” (Burton and Romney 1975:400). This simple, compelling method for collecting data about what goes with what was introduced to anthropology by Michael Burton, who analyzed Brim’s data using multidimensional scaling and hierarchical clustering. These powerful tools were brand new at the time and are used today across the social sciences (Burton 1968, 1972). (We’ll get back to MDS and clustering in chapter 21 on multivariate analysis.)

Informants often ask two questions when asked to do a pile sort: (1) “What do you mean by ‘belong together’?” and (2) “Can I put something in more than one pile?” The answer to the first question is “There are no right or wrong answers. We want to learn what *you* think about these things.”

The easy answer to the second question is “no,” because there is one card per item and a card can only be in one pile at a time. This answer cuts off a lot of information, however, because people can think of items in a cultural domain along several dimensions at once. For example, in a pile sort of consumer electronics, someone might want to put a DVD recorder in one pile with TVs (for the obvious association) and in another pile with camcorders (for another obvious association), but might not want to put camcorders and TVs

in the same pile. One way to handle this problem is to have duplicate cards that you give to people when they want to put an item into more than one pile, but be warned that this can complicate analysis of the data. An alternative is to ask informants to do **multiple free pile sorts** of the same set of objects.

The P-3 Game

In a series of papers, John Roberts and his coworkers used pile sorts and rating tasks to study how people perceive various kinds of behaviors in games (see, for example, Roberts and Chick 1979; Roberts and Nattress 1980). One “game,” studied by Roberts et al. (1980) is pretty serious: searching for foreign submarines in a P-3 airplane. The P-3 is a four-engine, turboprop, low-wing aircraft that can stay in the air for a long time and cover large patches of ocean. It is also used for search-and-rescue missions. Making errors in flying the P-3 can result in career damage and embarrassment, at least, and injury or death, at worst.

Through extensive, unstructured interviews with Navy P-3 pilots, Roberts et al. isolated 60 named flying errors. (This is the equivalent of extracting a free list from your interviews.) Here are a few of the errors: flying into a known thunderstorm area; taking off with the trim tabs set improperly; allowing the prop wash to cause damage to other aircraft; inducing an autofeather by rapid movement of power level controls. Roberts et al. asked 52 pilots to do a free pile sort of the 60 errors and to rate each error on a 7-point scale of “seriousness.”

They also asked the pilots to rank a subset of 13 errors on four criteria: (1) how much each error would “rattle” a pilot; (2) how badly each error would damage a pilot’s career; (3) how embarrassing each error would be to commit; and (4) how much “fun” it would be to commit each error. Flying into a thunderstorm on purpose, for example, could be very damaging to a pilot’s career, and extremely embarrassing if he had to abort the mission and turn back in the middle (when Roberts et al. did their research in the 1970s, all P-3 pilots were men). But if the mission was successful, then taking the risk of committing a very dangerous error would be a lot of fun for pilots who are, as Roberts called them, “high self-testers” (personal communication).

Inexperienced pilots rated “inducing an autofeather” as more serious than did highly experienced pilots. Inducing an autofeather is more embarrassing than it is dangerous and it’s the sort of error that experienced pilots just don’t make. On the other hand, as the number of air hours increased, so did pilots’ view of the seriousness of “failure to use all available navigational aids to determine position.” Roberts et al. suggested that inexperienced pilots might not have had enough training to assess the seriousness of this error correctly.

The Lumper-Splitter Problem

Most investigators use the **free pile sort** method (also called the **unconstrained pile sort method**), where people are told that they can make as many piles as they like, so long as they don't make a separate pile for each item or lump all the items into one pile. Like the triad test, the free pile sort presents people with a common set of stimuli, but there's a crucial difference: with free pile sorts, people can group the items together as they see fit. The result is that some people will make many piles, others will make few, and this causes the **lumper-splitter** problem (Weller and Romney 1988:22).

In a pile sort of animals, for example, some informants will put all the following together: giraffe, elephant, rhinoceros, zebra, wildebeest. They'll explain that these are the "African animals." Others will put giraffe, elephant, and rhino in one pile, and the zebra and wildebeest in another, explaining that one is the "large African animal" pile and the other is the "medium-sized African animal pile."

While they can't put *every* item in its own pile, lots of people put *some* items in singleton piles, explaining that each item is unique and doesn't go with the others. It's fine to ask informants why they made each pile of items, but wait until they finish the sorting task so you don't interfere with their concentration. And don't hover over informants. Find an excuse to walk away for a couple of minutes after they get the hang of it.

Since triad tests present each respondent with exactly the same stimuli, you can compare the data across individuals. Free pile sorts tell you what the structure of the data looks like for a group of people—sort of group cognition—but you can't compare the data from individuals. On the other hand, with pile sorts, you can have as many as 50 or 60 items. As with any measurement, each of these methods—triad tests and free pile sorts—has its advantages and disadvantages.

Pile Sorts with Objects

Pile sorts can also be done with objects. James Boster (1987) studied the structure of the domain of birds among the Aguaruna Jívaro of Peru. He paid people to bring him specimens of birds and he had the birds stuffed. He built a huge table out in the open, laid the birds on the table, and asked the Aguaruna to sort the birds into groups.

Carl Kendall led a team project in El Progreso, Honduras, to study beliefs about dengue fever (Kendall et al. 1990). Part of their study involved a pile sort of the nine most common flying insects in the region. They mounted specimens of the insects in little boxes and asked people to group the insects in

terms of “those that are similar.” Some fieldworkers have used photographs of objects as stimuli for a pile sort.

Borgatti (1999:133) points out that physical stimuli, like images or objects, make people focus on form rather than function. In fact, when asked to sort drawings of fish, fishermen in North Carolina sorted on shape—the long thin ones, the ones with a big dorsal fin, the small roundish ones (Boster and Johnson 1989). “In contrast,” says Borgatti (1999:133), “sorting *names* of fish allows hidden attributes to affect the sorting”—things like taste or how much of a struggle fish put up. “If you are after shared cultural beliefs,” says Borgatti, “I recommend keeping the stimulus as abstract as possible” (1992b:6).

Pile Sorts and Taxonomic Trees

Pile sorting is an efficient method for generating **taxonomic trees** (Werner and Fenton 1973). Simply hand informants the familiar pack of cards, each of which contains some term in a cultural domain. Informants sort the cards into piles, according to whatever criterion makes sense to them. After the first sorting, informants are handed each pile and asked to go through the exercise again. They keep doing this until they say that they cannot subdivide piles any further. At each sorting level, informants are asked if there is a word or phrase that describes each pile.

Perchonock and Werner (1969) used this technique in their study of Navajo animal categories. After an informant finished doing a pile sort of animal terms, Perchonock and Werner built a branching tree diagram (such as that shown in figure 11.1) from the data. They would ask the informant to make up sentences or phrases that expressed some relation between the nodes. They found that informants intuitively grasped the idea of tree representations for taxonomies. (For more about folk taxonomies, see chapter 18.)

Pile Sorts and Networks

I’ve used pile sorts to study the social structure of institutions such as prisons, ships at sea, and bureaucracies, and also to map the cognitively defined social organization of small communities. I simply hand people a deck of cards, each of which contains the name of one of the people in the group, and ask informants to sort the cards into piles, according to their own criteria.

The results tell me how people in the various components of an organization (managers, production workers, advertising people; or guards, counselors, prisoners; or seamen, deck officers, engine room personnel; or men and women in a small Greek village) think about the social structure of the group. Instead of “what goes with what,” I learn “who goes with whom.” Then I ask

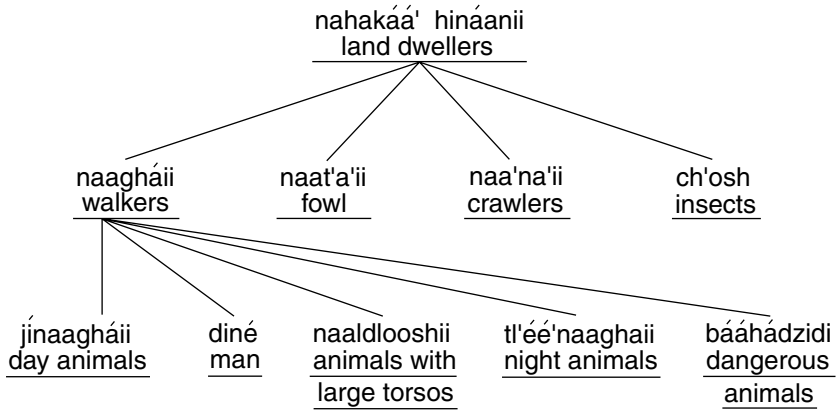


Figure 11.1. Part of the Navajo animal kingdom, derived from a pile sort.
 SOURCE: N. Perchonock and O. Werner, "Navaho Systems of Classification: Some Implications for Ethnoscience," *Ethnology*, Vol. 8, pp. 229–42. Copyright © 1969. Reprinted with permission.

informants to explain *why* people appear in the same pile. This produces a wealth of information about the cognitively defined social structure of a group.

Paired Comparisons

The method of paired comparisons is an alternative way to get rank orderings of a list of items in a domain. Remember, for any set of things, there are $n(n-1)/2$ pairs of those things. Suppose you have a list of five colors: red, green, yellow, blue, and brown. Figure 11.2 shows the paired comparison test to find out an informant's rank-ordered preference for these five colors.

You might say: "Here are two animals. Which one is the more _____?" where the blank is filled in by "vicious," or "wild," or "smarter," or some other descriptor.

You could ask informants to choose the "illness in this pair that is more life threatening," or "the food in this pair that is better for you," or "the crime in this pair that you're most afraid of."

I've presented the pairs in figure 11.2 in such a way that you can easily see how the 10 of them exhausts the possibilities for five items. When you present a paired comparison test to an informant, be sure to scramble the order of the pairs to ensure against **order effects**—that is, where something about the order of the items in a list influences the choices that informants make. You can use *Anthropac* to do this.

To find the rank order of the list for each informant, you simply count up how many times each item in a list "wins"—that is, how many times it was

In each of the following pairs of colors, please circle the one you like best:

RED	GREEN
RED	YELLOW
RED	BLUE
RED	BROWN
GREEN	YELLOW
GREEN	BLUE
GREEN	BROWN
YELLOW	BLUE
YELLOW	BROWN
BLUE	BROWN

Figure 11.2. A paired comparison test.

circled. If you are studying illnesses and cancer is on the list, and the question is “which of these pairs of illnesses is more life threatening,” you expect to find it circled each time it is paired with another illness—except, perhaps, when it is paired with AIDS. Since this is so predictable, it’s not very interesting. It gets really interesting when you have illnesses like diabetes and high blood pressure in your list and you compare the average rank ordering among various ethnic groups.

The paired comparison technique has a lot going for it. People make one judgment at a time, so it’s much easier on them than asking them to rank order a list of items by staring at all the items at once. Also, you can use paired comparisons with nonliterate informants by reading the list of pairs to them, one at a time, and recording their answers.

Like triad tests, paired comparisons can only be used with a relatively short list of items in a domain. With 20 items, for example, informants have to make 190 judgments.

Rankings and Ratings

Rank ordering produces interval-level data, though not all behaviors or concepts are easy to rank. Hammel (1962) asked people in a Peruvian village to rank order the people they knew in terms of prestige. By comparing the lists from different informants, Hammel was able to determine that the men he tested all had a similar view of the social hierarchy. Occupations can easily be rank ordered on the basis of prestige, or lucrateness.

Or even accessibility. The instructions to respondents would be “Here is a

list of occupations. Please rank them in order, from most likely to least likely that your daughter will have this occupation.” Then ask respondents to do the same thing for their sons. (Be sure to assign people randomly to doing the task for sons or daughters first.) Then compare the average ranking of accessibility against some independent variables and test for intracultural differences among ethnic groups, genders, age groups, and income groups.

Weller and Dungy (1986) studied breast-feeding among Hispanic and Anglo women in southern California. They asked 55 informants for a free list of positive and negative aspects of breast- and bottle-feeding. Then they selected the 20 most frequently mentioned items in this domain and converted the items to neutral, similarly worded statements. A few examples: “A way that doesn’t tie you down, so you are free to do more things”; “A way that your baby feels full and satisfied”; “A way that allows you to feel closer to your baby.”

Next, Weller and Dungy asked 195 women to rank the 20 statements. The women were asked which statement was most important to them in selecting a method of feeding their baby; which was the next most important to them; and so on. In the analysis, Weller and Dungy were able to relate the average rank order for Hispanics and for Anglos to independent variables like age and education.

Everyone is familiar with **rating scales**—all those agree-disagree, approve-disapprove instruments that populate the surveys we’ve been filling out all our lives. Rating scales are powerful data generators. They are so powerful and so ubiquitous that they deserve a whole chapter—which comes up next.

12



Scales and Scaling

This chapter is about building and using **composite measures**. I'll cover four kinds of composite measures: (1) indexes, (2) Guttman scales, (3) Likert scales, and (4) semantic differential scales. These four are the most commonly used in social research today. A fifth, magnitude scaling, is less common, but it's very interesting and it will give you an idea of the clever things that are going on in the field of scaling these days. First, though, some basic concepts of scaling.

Simple Scales: Single Indicators

A scale is a device for assigning units of analysis to categories of a variable. The assignment is usually done with numbers, and questions are used a lot as scaling devices. Here are three typical scaling questions:

1. "How old are you?"

You can use this question to assign individuals to categories of the variable "age." In other words, you can *scale* people by age. The number that this first question produces has ratio properties (someone who is 50 is twice as old as someone who is 25).

2. "How satisfied are you with your classes this semester? Are you satisfied, neutral, or unsatisfied?"

You can use *this* question to assign people to one of three categories of the variable "satisfied." That is, you can *scale* them according to how satisfied they are with their classes. Suppose we let satisfied = 3, neutral = 2, and unsatisfied = 1. Someone who is assigned the number 3 is *more* satisfied than someone who is assigned the number 1. We don't know if that means 3 times more satisfied, or

10 times, or just marginally more satisfied, so this scaling device produces numbers that have ordinal properties.

3. “Do you consider yourself to be Protestant, Catholic, Jewish, Muslim, some other religion? Or do you consider yourself as having no religion?”

This scaling device lets you assign individuals to—that is, *scale them* by—categories of the variable “religious affiliation.” Let Protestant = 1, Catholic = 2, Jewish = 3, Muslim = 4, and no religion = 5. The numbers produced by *this* device have nominal properties. You can’t add them up and find the average religion.

These three questions have different content (they tap different concepts), and they produce numbers with different properties, but they have two very important things in common: (1) all three questions are devices for scaling people and (2) in all three cases, the respondent is the principal source of measurement error.

When you use your own judgment to assign units of analysis to categories of a scaling device, *you* are the major source of measurement error. In other words, if you assign individuals by your own observation to the category “male” or “female,” then any mistakes you make in that assignment (in scaling people by sex) are *yours*.

Complex Scales: Multiple Indicators

So, a single question on a questionnaire is technically a scale if it lets you assign the people you’re studying to categories of a variable. A lot of really interesting variables however, are complex and can’t easily be assessed with single indicators. What single question could you ask an ethnic Chinese shopkeeper in Jakarta to measure how assimilated they were to Indonesian national culture? Could you measure the amount of stress people are experiencing by asking them a single question? We try to measure complex variables like these with complex instruments—that is, instruments that are made up of several indicators. These complex instruments are what people commonly call scales.

A classic concept in all of social research is “socioeconomic status,” or SES. Sociologists and psychologists often measure it in the industrialized countries by combining measures of income, education, and occupational prestige. Each of these measures is, by itself, an operationalization of the concept SES, but none of the measures captures the complexity of the idea of “socioeconomic status.” Each indicator captures a piece of the concept, and together the indicators produce a single measurement of SES. (See Ensminger and Fothergill [2003] and Oakes and Rossi [2003] for more on measuring SES.)

Of course, some variables are best measured by single indicators and, by Ockham's razor, we would never use a complex scale to measure something when a simple scale will do. So: The function of **single-indicator scales** is to assign units of analysis to categories of a variable. The function of **composite measures**, or complex scales, is exactly the same, but they are used when single indicators won't do the job.

Indexes

The most common composite measure is a **cumulative index**. Indexes are made up of several items, all of which count the same. Indexes are everywhere. The Dow-Jones Industrial Average is an index of the prices of 30 stocks that are traded on the New York Stock Exchange. The U.S. Consumer Price Index is a measure of how much it costs to buy a fixed set of consumer items in the United States. We use indexes to measure people's health risks: the risk of contracting HIV, of getting lung cancer, of having a heart attack, of giving birth to an underweight baby, of becoming an alcoholic, of suffering from depression, and on and on.

And, of course, we use indexes with a vengeance to measure cognitive and physical functions. Children in the industrial societies of the world begin taking intelligence tests, achievement tests, and tests of physical fitness from the first day they enter school—or even before that. Achievement indexes—like the SAT, ACT, and GRE—affect so many people in the United States, there's a thriving industry devoted to helping children and adolescents do well on these tests.

Indexes can be **criterion referenced** or **norm referenced**. If you've ever taken a test where the only way to get an "A" was to get at least 90%, you've had your knowledge of some subject assessed by a criterion-referenced index. If you've ever taken a test where getting an "A" required that you score in the top 10% of the class—even if the highest grade in the class were 70%—then you've had your knowledge of some subject assessed by a norm-referenced index.

Standardized tests (whether of achievement, or of performance, or of personality traits) are usually norm referenced: Your score is compared to the norms that have been established by thousands of people who took the test before you.

How Indexes Work

Multiple-choice exams are cumulative indexes. The idea is that asking just one question about the material in a course would not be a good indicator of

students' knowledge of the material. Instead, students typically are asked a bunch of multiple-choice questions.

Taken together, the reasoning goes, all the questions measure how well a student has mastered a body of material. If you take a test that has 60 multiple-choice questions and you get 45 correct, you get 45 points, one for each correct answer. That number, 45 (or 75%), is a cumulative index of how well you did on the test.

Note that in a cumulative index, it makes no difference *which* items are assigned to you. In a test of just 10 questions, for example, there are obviously just 10 ways to get one right—but there are 45 ways to get two right, 120 ways to get three right. . . . Students can get the same score of 80% on a test of 100 questions and miss entirely different sets of 20 questions. This makes cumulative indexes **robust**; they provide many ways to get at an underlying variable (in the case of an exam, the underlying variable is knowledge of the material).

On the other hand, stringing together a series of items to form an index doesn't guarantee that the composite measure will be useful—any more than stringing together a series of multiple-choice questions will fairly assess a student's knowledge of, say, anthropology.

We pretend that: (1) Knowledge is a **unidimensional** variable; (2) A fair set of questions is chosen to represent knowledge of some subject; and, therefore (3) A cumulative index is a fair test of the knowledge of that subject. We know that the system is imperfect, but we pretend in order to get on with life.

We don't have to pretend. When it comes to scaling units of analysis on complex constructs—like scaling countries on the construct of freedom or people on the construct of political conservatism—we can test the unidimensionality of an index with a technique called **Guttman scaling**.

Guttman Scales

In a **Guttman scale**, as compared to a cumulative index, the measurements for the items have a *particular pattern indicating that the items measure a unidimensional variable*. To understand the pattern we're looking for, consider the following three questions.

1. How much is 124 plus 14?
2. How much is $1/2 + 1/3 + 1/5 + 2/11$?
3. If $3X = 133$, then how much is X ?

If you know the answer to question 3, you probably know the answer to questions 1 and 2. If you know the answer to question 2, but not to 3, it's still

safe to assume that you know the answer to question 1. This means that, in general, *knowledge about basic math* is a unidimensional variable.

Now consider a highland, Aymara-speaking Bolivian village. As part of your study, you need to measure the level of acculturation of each person. That is, you want to assign a single number to each person—a number that represents how acculturated to nonindigenous, national Bolivian culture each person is. After some time in the community, you come to understand that there are three key points of acculturation: dress, language, and housing. As Indians acculturate, they dress in Western clothes, learn to speak Spanish fluently (in addition to or instead of Aymara), and build Western-style houses.

From your ethnographic work, you reason that people need significant wealth to afford a Western-style house, with all the imported materials that building one entails. People who have wealth participate in the national economy, which means that they must be fluent in Spanish. Anyone, however, can afford to adopt Western-style clothes, especially used clothing. According to your theory, Western dress is the easiest item to adopt; Spanish comes next; and then comes Western houses.

To test whether the indicators you've identified form a unidimensional, or Guttman scale, set up a table like table 12.1. It's not pretty.

Persons 1, 2, and 3 scored positive on all three items. They each get 3

TABLE 12.1
An Index That Scales with a Guttman Coefficient of Reproducibility <0.90

<i>Informant</i>	<i>Western clothes</i>	<i>Fluent Spanish</i>	<i>Western house</i>
1	+	+	+
2	+	+	+
3	+	+	+
4	+	+	—
5	+	+	—
6	+	+	—
7	+	—	—
8	—	—	—
9	—	—	—
10	—	+	—
11	—	+	—
12	—	+	—
13	—	—	+
14	—	—	+
15	—	—	+
16	+	—	+

points. The next three (4, 5, and 6) wear Western clothes and speak fluent Spanish, but live in indigenous-style houses. They each get 2 points. Person 7 wears Western clothes, but does not speak fluent Spanish, and does not live in a Western-style house. This informant gets 1 point on the acculturation index. Persons 8 and 9 have no acculturation points. They wear traditional dress, speak little Spanish, and live in traditional homes. So far so good.

The next three (10, 11, 12) speak fluent Spanish but wear traditional dress and live in traditional houses. The next three (13, 14, 15) live in Western-style homes but wear traditional dress and are not fluent in Spanish. Finally, person 16 wears Western clothes and lives in a Western house, but is not fluent in Spanish.

If we had data from only the first nine respondents, the data would form a perfect Guttman scale. For those first nine respondents, in other words, the three behaviors are indicators of a unidimensional variable, acculturation.

The Coefficient of Reproducibility

Unfortunately, we've got those other seven people to deal with. For whatever reasons, informants 10–16 do not conform to the pattern produced by the data from informants 1–9. The data for persons 10–16 are “errors” in the sense that their data diminish the extent to which the index of alienation forms a perfect scale. To test how closely any set of index data reproduces a perfect scale, apply Guttman's **coefficient of reproducibility**, or CR. The formula for Guttman's CR is:

$$1 - \frac{\text{number of errors}}{\text{number of entries}} \quad \textbf{Formula 12.1}$$

Given the pattern in table 12.1 (and from our hypothesis about the order in which people adopt the three indicators of acculturation), we don't expect to see those minus signs in column 1 for respondents 10, 11, and 12. If the data scaled according to our hypothesis, then anyone who speaks fluent Spanish and lives in a traditional house should wear Western-style clothes, as is the case with informants 4, 5, and 6. *Those* informants have a score of 2. It would take three corrections to make cases 10, 11, and 12 conform to the hypothesis (you'd have to replace the minus signs in column one with pluses for respondents 10, 11, and 12), so we count cases 10, 11, and 12 as having one error each.

We don't expect to see the plus signs in column 3 for informants 13, 14, and 15. If our hypothesis were correct, anyone who has a plus in column 3 should have all pluses and a score of 3 on acculturation. If we give respondents 13, 14, and 15 a scale score of 3 (for living in a Western-style house), then

those three cases would be responsible for *six* errors—you'd have to stick in two pluses for each of the cases to make them come out according to the hypothesis. Yes, you could make it just three, not six errors, by sticking a minus sign in column 3. Some researchers use this scoring method, but I prefer the more conservative method of scoring more errors. It keeps you on your toes.

Finally, we don't expect that minus sign in column 2 of respondent 16's data. That case creates just one error (you only need to put in one plus to make it come out right). All together, that makes $3 + 6 + 1 = 10$ errors in the attempt to reproduce a perfect scale. For table 12.1, the CR is

$$1 - (10/48) = .79$$

which is to say that the data come within 21% of scaling perfectly. By convention, a coefficient of reproducibility of .90 or greater is accepted as a significant approximation of a perfect scale (Guttman 1950). I'm willing to settle for around .85, especially with the conservative method for scoring errors, but .79 just isn't up to it, so these data fail the Guttman test for unidimensionality.

Some Examples of a Guttman Scale

Robert Carneiro (1962, 1970) had an idea that cultural evolution is orderly and cumulative. If he is right, then cultures evolve by adding certain traits in an orderly way and should show a Guttman-scale-like pattern. Carneiro coded 100 cultures for 354 traits and looked at the pattern. Table 12.2 shows a sample of 12 societies and 11 traits.

When you collect data on cases, you don't know what (if any) pattern will emerge, so you pretty much grab cases and code them for traits in random order. The 12 societies and traits in Table 12.2a are in random order.

The first thing to do is arrange the pluses and minuses in their "best" possible order—the order that conforms most to the perfect Guttman scale—and compute the CR. We look for the trait that occurs most frequently (the one with the most pluses across the row) and place that one at the bottom of the matrix. The most frequently occurring trait is the existence of special religious practitioners. Then we look for the next most frequent trait and put it on the next to the bottom row of the matrix.

We keep doing this until we rearrange the data to take advantage of whatever underlying pattern is hiding in the matrix. The best arrangement of the pluses and minuses is shown in table 12.2b. Now we can count up the "errors" in the matrix and compute Guttman's coefficient of reproducibility. For these 12 societies and 11 traits, the coefficient is a perfect 1.0.

Of course, it's one thing to find this kind of blatant pattern in a matrix of

TABLE 12.2a
 Carneiro's Matrix Showing the Presence (+) or Absence (-) of 11 Culture Traits among 12 Societies. The Order of Both the Traits and the Societies Is Random.

<i>Society</i>	1	2	3	4	5	6	7	8	9	10	11	12
Political leader has considerable authority	+	+	-	-	+	-	-	+	-	+	-	-
Sumptuary laws	-	-	-	-	+	-	-	+	-	-	-	-
Headman, chief, or king	+	+	-	-	+	+	+	+	-	+	+	+
Surplus of food regularly produced	+	+	-	-	+	-	-	+	-	+	-	+
Trade between communities	+	+	+	+	+	+	+	+	-	+	+	+
Ruler grants audiences	-	+	-	-	+	-	-	+	-	+	-	-
Special religious practitioners	+	-	+	+	+	+	+	+	+	+	+	+
Paved streets	-	-	-	-	-	-	-	+	-	-	-	-
Agriculture provides $\geq 75\%$ of subsistence	+	+	-	-	+	-	+	+	-	+	-	+
Full-time service specialists	-	+	-	-	+	-	-	+	-	-	-	-
Settlements ≥ 100 persons	+	+	-	-	+	+	+	+	-	+	-	+

Societies: 1 Iroquois, 2 Marquesans, 3 Tasmanians, 4 Yahgan, 5 Dahomey, 6, Mundurucú, 7 Ao Naga, 8 Inca, 9 Semang, 10 Tanala, 11 Vedda, 12 Bontoc

SOURCE: *A Handbook of Method in Cultural Anthropology* by Raoul Naroll and Ronald Cohen, eds., copyright © 1970 by Raoul Naroll and Ronald Cohen. Used by permission of Doubleday, a division of Random House, Inc.

TABLE 12.2b
 The Data in Table 12.2a Rearranged: The Data Form a Perfect Guttman Scale

<i>Society</i>	1	2	3	4	5	6	7	8	9	10	11	12
Paved streets	-	-	-	-	-	-	-	-	-	-	-	+
Sumptuary laws	-	-	-	-	-	-	-	-	-	-	+	+
Full-time service specialists	-	-	-	-	-	-	-	-	-	+	+	+
Ruler grants audiences	-	-	-	-	-	-	-	-	+	+	+	+
Political leader has considerable authority	-	-	-	-	-	-	-	+	+	+	+	+
Surplus of food regularly produced	-	-	-	-	-	-	+	+	+	+	+	+
Agriculture provides $\geq 75\%$ of subsistence	-	-	-	-	-	+	+	+	+	+	+	+
Settlements ≥ 100 persons	-	-	-	-	+	+	+	+	+	+	+	+
Headman, chief, or king	-	-	-	+	+	+	+	+	+	+	+	+
Trade between communities	-	-	+	+	+	+	+	+	+	+	+	+
Special religious practitioners	-	+	+	+	+	+	+	+	+	+	+	+

Societies: 1 Tasmanians, 2 Semang, 3 Yahgan, 4 Vedda, 5 Mundurucú, 6 Ao Naga, 7 Bontoc, 8 Iroquois, 9 Tanala, 10 Marquesans, 11 Dahomey, 12 Inca

12 societies and 11 traits. Carneiro, you'll recall, coded 100 societies for 354 traits and then went looking for subsets of the data that showed the desired pattern. When he did this in the 1960s, it was heroic work. Today, *Anthropac* (Borgatti 1992a, 1992b) has a routine for looking at big matrices of pluses and minuses, rearranging the entries into the best pattern, calculating the CR, and showing you which units of analysis and traits to drop in order to find the optimal solution to the problem.

Data Scale, Variables Don't

DeWalt (1979) used Guttman scaling to test his index of material style of life in a Mexican farming community. He scored 54 informants on whether they possessed eight material items (a radio, a stove, a sewing machine, etc.) and achieved a remarkable CR of .95. This means that, *for his data*, the index of material style of life is highly reliable and differentiates among informants.

Remember: *Only data scale, not variables*. If the items in a cumulative index form a Guttman scale with 0.90 CR or better, we can say that, *for the sample we've tested*, the concept measured by the index is unidimensional. That is, the items are a composite measure of one and only one underlying concept. DeWalt's data show that, for the informants he studied, the concept of "material style of life" is unidimensional—at least for the indicators he used.

The Guttman technique tests whether unidimensionality holds for a particular set of data. An index must be checked for its Guttman scalability each time it is used on a population. My hunch is that DeWalt's material-style-of-life scale has its analog in nearly all societies. The particular list of items that DeWalt used in rural Mexico may not scale in a middle-class neighborhood of Ulan Bator, but *some* list of material items *will* scale there. You just have to find them.

The way to do this is to code every household in your study for the presence or absence of a list of material items. The particular list could emerge from participant observation or from informal interviews. Then you'd use *Anthropac* to sort out the matrix, drop some material items, and build the material index that has a CR of 0.90 or better. Greg Guest (2000) did this in his study of 203 households in an Ecuadorian fishing village. He gave each household a score from 1 to 7, depending how many material items they had. That score correlated significantly with the education level of the head of each household. (We'll get to correlation and statistical significance in chapter 20.) Since we expect a correlation between wealth and education, this adds construct validity to Guest's scale.

Careful, though. Oliver Kortendick tried to develop a Guttman scale of

wealth in village in Papua New Guinea. The idea of property ownership may not have existed in that culture prior to contact with Europeans and Australians in the mid-20th century. It was well understood when Kortendick got there, but some things, like cars, were too expensive for anyone there to possess on their own. So villagers bought and owned those items collectively (Kortendick, personal communication).

Indexes That Don't Scale

Indexes that do not scale can still be useful in comparing populations. Dennis Werner (1985) studied psychosomatic stress among Brazilian farmers who were facing the uncertainty of having their lands flooded by a major dam. He used a 20-item stress index developed by Berry (1976).

Since the index did not constitute a unidimensional scale, Werner could not differentiate among his *informants* (in terms of the amount of stress they were under) as precisely as DeWalt could differentiate among *his* informants (in terms of their quality of life). But farmers in Werner's sample gave a stress response to an average of 9.13 questions on the 20-item test, while Berry had found that Canadian farmers gave stress responses on an average of 1.79 questions. It is very unlikely that a difference of such magnitude between two *populations* would occur by chance.

Likert Scales

Perhaps the most commonly used form of scaling is attributed to Rensis Likert (1932). Likert introduced the ever-popular 5-point scale that we talked about in chapter 10 (on questionnaire construction). Recall that a typical question might read as follows:

Please consider the following statements carefully. After each statement, circle the answer that most reflects your opinion. Would you say you agree a lot with the statement, agree a little, are neutral, disagree a little, or disagree a lot with each statement? Ok, here's the first statement:

When I need credit to bring my bananas to market, I can just go to the agricultural bank in Ralundat and they give it to me.

Agree a lot _____

Agree _____

Neutral _____

Disagree a little _____

Disagree a lot _____

The 5-point scale might become 3 points or 7 points, and the agree-disagree scale may become approve-disapprove, favor-oppose, or excellent-bad, but the principle is the same. These are all **Likert-type** scales.

I say “Likert-type scales” rather than just “Likert scales” because Likert did more than just introduce a format. He was interested in measuring internal states of people (attitudes, emotions, orientations) and he realized that most internal states are multidimensional. You hear a lot of talk these days about conservatives and liberals, but the concept of political orientation is very complex. A person who is liberal on matters of domestic policy—favoring government-supported health care, for example—may be conservative on matters of foreign political policy—against involvement in any foreign military actions. Someone who is liberal on matters of foreign economic policy—favoring economic aid for all democracies that ask for it—may be conservative on matters of personal behavior—against same-sex marriage, for example.

The liberal-conservative dimension on matters of personal behavior is also complicated. There’s no way to assign people to a category of this variable by asking one question. People can have live-and-let-live attitudes about sexual preference and extramarital sex and be against a woman’s right to an abortion on demand.

Of course, there are packaging effects. People who are conservative on one dimension of political orientation are *likely* to be conservative on other dimensions, and people who are liberal on one kind of personal behavior are *likely* to be liberal on others. Still, no single question lets you scale people in general on a variable as complex as “attitude toward personal behavior,” let alone “political orientation.” That’s why we need composite scales.

Steps in Building a Likert Scale

Likert’s method was to take a long list of possible scaling items for a concept and find the subsets that measured the various dimensions. If the concept were unidimensional, then one subset would do. If it were multidimensional, then several subsets would be needed. Here are the steps in building and testing a Likert scale.

1. Identify and label the variable you want to measure. This is generally done by induction—that is, from your own experience (Spector 1992:13). After you work in some area of research for a while, you’ll develop some ideas about the variables you want to measure. The people you talk to in focus groups, for example, may impress you with the idea that “people are afraid of crime around here,” and you decide to scale people on the variable “fear of crime.”

You may observe that some people seem to have a black belt in shopping, while others would rather have root canal surgery than set foot in a mall. The task is then to scale (measure) people on a variable you might call “shopping orientation,” with all its multidimensionality. You may need a subscale for “shopping while on vacation,” another for “car shopping,” and another for “shopping for clothing that I really need.” (The other way to identify variables is by deduction. This generally involves analyzing similarity matrices, about which more in chapters 16 and 21.)

2. Write a long list of indicator questions or statements. This is usually another exercise in induction. Ideas for the indicators can come from reading the literature on whatever research problem has captured you, from personal experience, from ethnography, from reading newspapers, from interviews with experts.

Free lists are a particularly good way to get at indicators for some variables. If you want to build a scaling device for the concept of “attitudes toward growing old,” you could start by asking a large group of people to “list things that you associate with growing old” and then you could build the questions or statements in a Likert scale around the items in the list.

Be sure to use both negative and positive indicators. If you have a statement like “Life in Xakalornga has improved since the missionaries came,” then you need a negatively worded statement for balance like “The missionaries have caused a lot of problems in our community.” People who agree with positive statements about missionaries should disagree with negative ones.

And don’t make the indicator items extreme. Here’s a badly worded item: “The coming of the missionaries is the most terrible thing that has ever happened here.” Let people tell *you* where they stand by giving them a range of response choices (strongly agree–strongly disagree). Don’t bludgeon people with such strongly worded scale items that they feel forced to reduce the strength of their response.

In wording items, all the cautions from chapter 10 on questionnaire design apply: Remember who your respondents are and use *their* language. Make the items as short and as uncomplicated as possible. No double negatives. No double-barreled items. Here is a terrible item:

On a scale of 1 to 5, how much do you agree or disagree with the following statement: “Everyone should speak Hindi and give up their tribal language.”

People can agree or disagree with both parts of this statement, or agree with one part and disagree with the other.

When you get through, you should have four or five times the number of

items as you think you'll need in your final scale. If you want a scale of, say, six items, use 25 or 30 items in the first test (DeVellis 1991:57).

3. Determine the type and number of response categories. Some popular response categories are agree-disagree, favor-oppose, helpful-not helpful, many-none, like me-not like me, true-untrue, suitable-unsuitable, always-never, and so on. Most Likert scale items have an odd number of response choices: three, five, or seven. The idea is to give people a range of choices that includes a midpoint. The midpoint usually carries the idea of neutrality—neither agree nor disagree, for example. An even number of response choices forces informants to “take a stand,” while an odd number of choices lets informants “sit on the fence.”

There is no best format. But if you ever want to combine responses into just two categories (yes-no, agree-disagree, like me-not like me), then it's better to have an even number of choices. Otherwise, you have to decide whether the neutral responses get collapsed with the positive answers or the negative answers—or thrown out as missing data.

4. Test your item pool on some respondents. Ideally, you need at least 100—or even 200—respondents to test an initial pool of items (Spector 1992:29). This will ensure that: (1) You capture the full variation in responses to all your items; and (2) The response variability represents the variability in the general population to which you eventually want to apply your scale.
5. Conduct an **item analysis** to find the items that form a unidimensional scale of the variable you're trying to measure. More on item analysis coming up next.
6. Use your scale in your study and run the item analysis again to make sure that the scale is holding up. If the scale does hold up, then look for relations between the scale scores and the scores of other variables for persons in your study.

Item Analysis

This is the key to building scales. The idea is to find out which, among the many items you're testing, need to be kept and which should be thrown away. The set of items that you keep should tap a single social or psychological dimension. In other words, the scale should be unidimensional.

In the next few pages, I'm going to walk through the logic of building scales that are unidimensional. Read these pages very carefully. At the end of this section, I'll advocate using **factor analysis** to do the item analysis quickly, easily, and reliably. No fair, though, using factor analysis for scale construction until you understand the logic of scale construction itself.

There are three steps to doing an item analysis and finding a subset of items that constitute a unidimensional scale: (1) scoring the items, (2a) taking the

interitem correlation and (2b) **Cronbach's alpha**, and (3) taking the **item-total correlation**.

1. SCORING THE RESPONSES

The first thing to do is make sure that all the items are properly scored. Assume that we're trying to find items for a scale that measures the strength of support for training in research methods among anthropology students. Here are two potential scale items:

Training in statistics should be required for all undergraduate students of anthropology.

1	2	3	4	5
Strongly disagree	Disagree	Neutral	Agree	Strongly agree

Anthropology undergraduates don't need training in statistics.

1	2	3	4	5
Strongly disagree	Disagree	Neutral	Agree	Strongly agree

You can let the big and small numbers stand for any direction you want, but you must be consistent. Suppose we let the bigger numbers (4 and 5) represent support for training in statistics and let the smaller numbers (1 and 2) represent lack of support for that concept. Those who circle "strongly agree" on the first item get a 5 for that item. Those who circle "strongly agree" on the second item get scored as 1.

2A. TAKING THE INTERITEM CORRELATION

Next, test to see which items contribute to measuring the construct you're trying to get at, and which don't. This involves two calculations: the intercorrelation of the items and the correlation of the item scores with the total scores for each informant. Table 12.3 shows the scores for three people on three items, where the items are scored from 1 to 5.

To find the interitem correlation, we would look at all pairs of columns. There are three possible pairs of columns for a three-item matrix. These are shown in table 12.4.

A simple measure of how much these pairs of numbers are alike or unlike involves, first, adding up their *actual differences*, Σ_d , and then dividing this by the total *possible differences*, Max_d .

In the first pair, the actual difference between 1 and 3 is 2; the difference

TABLE 12.3
The Scores for Three People on Three Likert Scale Items

Person	Item		
	1	2	3
1	1	3	5
2	5	2	2
3	4	1	3

between 5 and 2 is 3; the difference between 4 and 1 is 3. The sum of the differences is $\sum_d 2 + 3 + 3 = 8$.

For each item, there could be as much as 4 points difference—in Pair 1, someone could have answered 1 to item 1 and 5 to item 2, for example. So for three items, the total possible difference, Max_d , would be $4 \times 3 = 12$. The actual *difference* is 8 out of a possible 12 points, so items 1 and 2 are $8/12 = 0.67$ *different*, which means that these two items are $1 - \sum_d/\text{Max}_d = .33$ *alike*. Items 1 and 3 are also 0.33 alike, and items 2 and 3 are 0.67 alike.

Items that measure the same underlying construct should be related to one another. If I answer “strongly agree” to the statement “Training in statistics should be required for all undergraduate students of anthropology,” then (if I’m consistent in my attitude and if the items that tap my attitude are properly worded) I should strongly disagree with the statement that “anthropology undergraduates don’t need training in statistics.” If everyone who answers “strongly agree” to the first statement answers “strongly disagree” to the second, then the items are perfectly correlated.

2B. CRONBACH’S ALPHA

Cronbach’s alpha is a statistical test of how well the items in a scale are correlated with one another. One of the methods for testing the unidimension-

TABLE 12.4
The Data from the Three Pairs of Items in Table 12.3

Pair 1		Diff	Pair 2		Diff	Pair 3		Diff
1	3	2	1	5	4	3	5	2
5	2	3	5	2	3	2	2	0
4	1	3	4	3	1	1	3	2
\sum_d (Sum of the diff’s)					8	4		
\sum_d / Max_d					0.67	0.33		
$1 - (\sum_d / \text{Max}_d)$					0.33	0.67		

ality of a scale is called the **split-half reliability** test. If a scale of, say, 10 items, were unidimensional, all the items would be measuring parts of the same underlying concept. In that case, any five items should produce scores that are more or less like the scores of any other five items. This is shown in table 12.5.

TABLE 12.5
The Schematic for the Split-Half Reliability Test

<i>Person</i>	<i>Split A: Score on items 1–5</i>	<i>Split B: Score on items 6–10</i>
1	X ₁	Y ₁
2	X ₂	Y ₂
3	X ₃	Y ₃
.	.	.
.	.	.
N	X _n	Y _n
	Total for A	Total for B

Split Halves and the Combinations Rule

There are many ways to split a group of items into halves and each split will give you a different set of totals. Here's the formula for selecting n elements from a set of N elements, paying no attention to the ordering of the elements:

$$\frac{N!}{n! (N-n)!} \quad \text{Formula 12.2}$$

If you have 10 respondents, then there are $10!/5!(10 - 5)! = 252$ ways to split them into halves of five each. For 20 items, there are 184,756 possible splits of 10 each. Cronbach's **coefficient alpha** provides a way to get the average of all these split-half calculations directly. The formula for Cronbach's alpha is:

$$\alpha = \frac{N_\rho}{1 + \rho (N-1)} \quad \text{Formula 12.3}$$

where ρ (the Greek letter *rho*) is the average interitem correlation—that is, the average correlation among all pairs of items being tested.

By convention, a good set of scale items should have a Cronbach's alpha of 0.80 or higher. Be warned, though, that if you have a long list of scale items, the chances are good of getting a high alpha coefficient. An interitem correla-

tion of just .14 produces an alpha of .80 in a set of 25 items (DeVellis 1991:92).

Eventually, you want an alpha coefficient of 0.80 or higher for a *short* list of items, all of which hang together and measure the same thing. Cronbach's alpha will tell you if your scale hangs together, but it won't tell you which items to throw away and which to keep. To do that, you need to identify the items that do not discriminate between people who score high and people who score low on the total set of items.

3. FINDING THE ITEM-TOTAL CORRELATION

First, find the total score for each person. Add up each respondent's scores across all the items. Table 12.6 shows what it would look like if you tested 50 items on 200 people (each x is a score for one person on one item).

TABLE 12.6
Finding the Item-Total Correlation

<i>Person</i>	<i>Item 1</i>	<i>Item 2</i>	<i>Item 3</i>	.	.	<i>Item 50</i>
1	x	x	x	.	.	x
2	x	x	x	.	.	x
3	x	x	x	.	.	x
.
.
200	x	x	x	.	.	x

For 50 items, scored from 1 to 5, each person could get a score as low as 50 (by getting a score of 1 on each item) or as high as 250 (by getting a score of 5 on each item). In practice, of course, each person in a survey will get a total score somewhere in between.

A rough-and-ready way to find the items that discriminate well among respondents is to divide the respondents into two groups, the 25% with the highest total scores and the 25% with the lowest total scores. Look for the items that the two groups have in common. Those items are *not discriminating* among informants with regard to the concept being tested. Items that fail, for example, to discriminate between people who strongly favor training in methods (the top 25%) and people who don't (the bottom 25%) are not good items for scaling people in this construct. Throw those items out.

There is a more formal way to find the items that discriminate well among respondents and the items that don't. This is the **item-total correlation**. Table 12.7 shows the data you need for this:

TABLE 12.7
The Data for the Interim Correlation

<i>Person</i>	<i>Total Score</i>	<i>Item 1</i>	<i>Item 2</i>	<i>Item 3</i>	.	.	<i>Item 50</i>
1	x	x	x	x	.	.	x
2	x	x	x	x	.	.	x
3	x	x	x	x	.	.	x
.
.
<i>N</i>	x	x	x	x	.	.	x

With 50 items, the total score gives you an idea of where each person stands on the concept you're trying to measure. If the interitem correlation were perfect, then every item would be contributing equally to our understanding of where each respondent stands. Of course, some items do better than others. The ones that don't contribute a lot will correlate poorly with the total score for each person. Keep the items that have the highest correlation with the total scores.

You can use any statistical analysis package to find the interitem correlations, Cronbach's alpha, and the item-total correlations for a set of preliminary scale items. Your goal is to get rid of items that detract from a high interitem correlation and to keep the alpha coefficient above 0.80. (For an excellent step-by-step explanation of item analysis, see Spector 1992:43–46.)

Testing for Unidimensionality with Factor Analysis

Factor analysis is a technique for data reduction. If you have 30 items in a pool of potential scale items, and responses from a sample of people to those pool items, factor analysis lets you reduce the 30 items to a smaller set—say, 5 or 6. Each item is given a score, called its **factor loading**. This tells you how much each item “belongs” to each of the underlying factors. (See chapter 21 for a brief introduction to factor analysis and Comrey [1992] for more coverage.)

If a scale is unidimensional, there will be a single, overwhelming factor that underlies all the variables (items) and all the items will “load high” on that single factor. If a scale is multidimensional, then there will be a series of factors that underlie sets of variables. Scale developers get a large pool of potential scale items (at least 40) and ask a lot of people (at least 200) to respond to the items. Then they run the factor analysis and select those items that load

high on the factor or factors (the underlying concept or concepts) they are trying to understand.

If you want to see what professional scale developers do, consult any of the following: Klonoff and Landrine (2000) (a scale for measuring acculturation among African Americans), Staats et al. (1996) (a scale measuring commitment to pets), Sin and Yau (2004) (a scale for measuring female role orientation in China), and Simpson and Gangstad (1991) (a scale that measures willingness to engage in uncommitted sexual relations).

Most anthropologists won't develop major scales for others to use, but what you *should* do is test the unidimensionality of any measures you develop for your own field data, using factor analysis—once you understand the principles of scale development that I've laid out here. (And just for the record, some anthropologists do develop scales. See Handwerker [1997] for a scale to measure family violence in Barbados and Gatz and Hurwicz [1990] for a scale to measure depression in old people.)

Semantic Differential Scales

I've always liked the semantic differential scaling method. It was developed in the 1950s by Charles Osgood and his associates at the University of Illinois and has become an important research tool in cognitive studies, including psychology, anthropology, and sociology (Osgood et al. 1957; Snider and Osgood 1969). It has also been used by thousands of researchers across the social sciences, and with good reason: The semantic differential test is easy to construct and easy to administer.

Osgood was interested in how people interpret things—inanimate things (like artifacts or monuments), animate things (like persons or the self), behaviors (like incest, or buying a new car, or shooting a deer), and intangible concepts (like gun control or literacy). Of course, this is exactly what Likert scales are designed to test, but instead of asking people to rate questionnaire items about things, Osgood tested people's feelings differently: He gave them a **target item** and a list of **paired adjectives** about the target. The adjective pairs could come from reading of the literature or from focus groups or from ethnographic interviews. Target items can be ideas (land reform, socialism, aggression), behaviors (smoking, running, hunting deer with a bow and arrow), objects (the mall, a courtroom, horses), environmental conditions (rain, drought, jungle) . . . almost anything.

Figure 12.1 is an example of a semantic differential test. The target is "having a cold." If you were taking this test right now, you'd be asked to place a check on each line, depending on your reaction to each pair of adjectives.

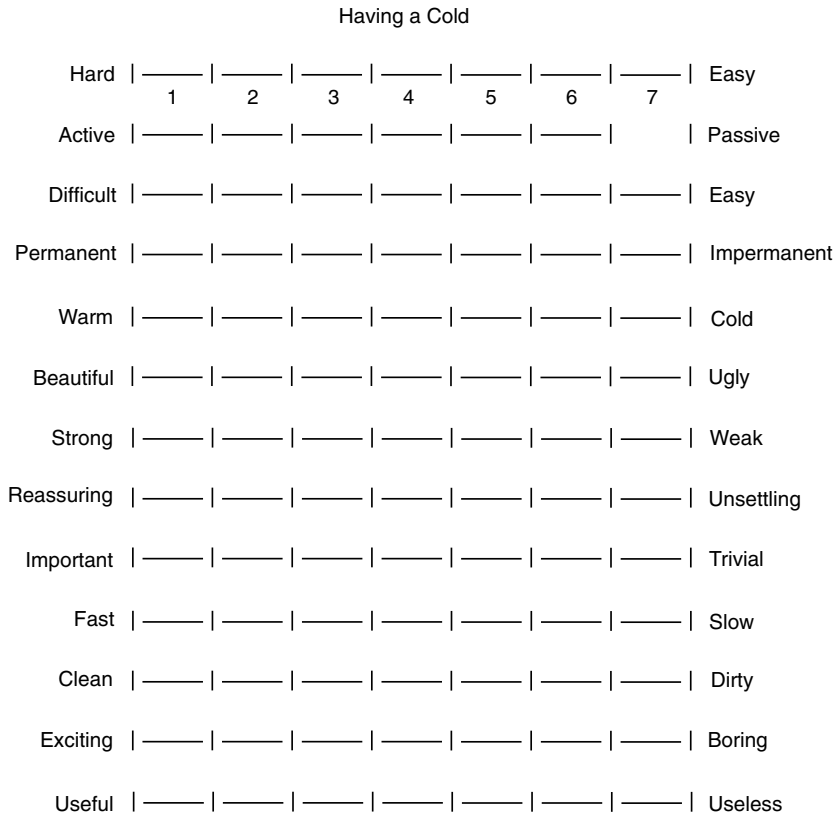


Figure 12.1. A semantic differential scale to test how people feel about having a cold. The dimensions in this scale are useful for measuring how people feel about many different things.

With a Likert scale, you ask people a series of questions that get at the target concept. In a semantic differential scale, you name the target concept and ask people to rate their feelings toward it on a series of variables. The semantic differential is usually a 7-point scale, as I've indicated in the first adjective pair above. (You can leave out the numbers and let people respond to just the visual form of the scale.) Your score on this test would be the sum of all your answers to the 14 adjective pairs.

Osgood and his associates did hundreds of replications of this test, using hundreds of adjective pairs, in 26 different cultures. Their analyses showed that in every culture, just three major kinds of adjectives account for most

of the variation in people's responses: **adjectives of evaluation** (good-bad, difficult-easy), **adjectives of potency** (strong-weak, dominant-submissive, etc.), and **adjectives of activity** (fast-slow, active-inactive, sedentary-mobile, etc.).

As the target changes, of course, you have to make sure that the adjective pairs make sense. The adjective pair ethical-corrupt works for some targets, but you probably wouldn't use it for having a cold.

Vincke et al. (2001) used the semantic differential scale to explore the meaning of 25 sex acts among gay men in Flanders, Belgium. Their informants scaled each act (anal insertive sex, anal receptive sex, insertive fellatio, receptive fellatio, interfemoral sex, and so on) on six paired dimensions: unsatisfying/satisfying, stimulating/dull, interesting/boring, emotional/unemotional, healthy/unhealthy, and safety/danger. Vincke et al. then compared results on the semantic differential for men who practiced safe sex (with one partner or with a condom) and men who practiced unsafe sex (multiple partners and without a condom) to see which sex acts were more gratifying for high-risk-taking and low-risk-taking men.

How Many Scale Choices?

We know from everyday experience that how we phrase a question in part determines the answer. We say things in ordinary conversation and, after listening to the response, we backtrack, fill in, and cover: "No, that's not what I meant. What I meant was. . . ." It won't come as a surprise, then, that our informants respond to the way we phrase our questions. They know that an interview is not an ordinary conversation, but it's a conversation nonetheless. They take cues from our questions, figure out what we want to know, and respond the best they can.

Tourangeau and Smith (1996) asked men and women the following question: "During the last 12 months, that is, since August/September 1993, how many men [women], if any, have you had intercourse with?" Some people were asked simply to tell the interviewer a number. Others were asked to choose one of the following: 0, 1, 2, 3, 4, 5 or more. And still others were asked to choose one of the following: 1-4, 5-9, 10-49, 50-99, 100 or more. It won't surprise you to learn that people report more sex partners when given high-end choices than when given low-end choices or an open-ended question (ibid.:292). Norbert Schwarz and his colleagues (1985) found the same thing when they asked people in Germany how many hours of television they watched. If you give people choices that *start* high, you get reports of more behavior.

Some Other Scales

The Cantril Ladder of Life

There are many interesting variations in the construction of scales. Hadley Cantril (1965) devised a 10-rung **ladder of life**, shown in figure 12.2. People

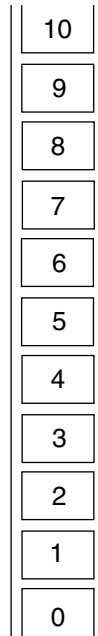


Figure 12.2. The ladder of life.

SOURCE: H. Cantril, "The Ladder of Life," *The Pattern of Human Concerns*. Copyright © 1965 by Rutgers, The State University. Reprinted by permission of Rutgers University Press.

are asked to list their concerns in life (financial success, healthy children, freedom from war, and so on). Then they are shown the ladder and are told that the bottom rung represents the worst-possible situation, while the top rung represents the best. For each of their concerns they are asked to point out where they are on the ladder right now, where they were 5 years ago, and where they think they'll be 5 years from now.

Note that the ladder of life is a **self-anchoring scale**. Respondents are asked to explain, in their own terms, what the top and bottom rungs of the ladder mean to them.

The ladder of life is a useful prop for interviewing nonliterate or semiliterate people. Hansen and McSpadden (1993), for example, used the technique

in their studies of Zambian and Ethiopian refugees in Zambia and the United States. In Zambia, Hansen actually constructed a small wooden ladder and found that the method worked well. McSpadden used several methods to explore how Ethiopian refugees adjusted to life in the United States. Even when other methods failed, McSpadden found that the ladder of life method got people to talk about their experiences, fears, and hopes (*ibid.*).

Keith et al. (1994) used a modified version of the ladder of life in their study of aging in seven cultures. In five of the sites (two in the United States, one in Hong Kong, and two in Ireland) where most informants were literate, they used a six-rung ladder. In Hong Kong, people were comfortable placing themselves *between* but not *on* rungs, so the team redesigned the ladder into a flight of stairs. Among the Herero and !Kung of Botswana, where many people were not literate, they replaced the ladder with the five fingers of the interviewer's hand (*ibid.*:xxx, 113).

Be careful to tell respondents exactly what you want when you use any kind of visual prop. Jones and Nies (1996) used Cantril's ladder to measure the importance of exercise to elderly African American women. Well, at least Jones and Nies *thought* that's what they were measuring. The mean for the ladder rating was about 9 on a scale of 1–10. Respondents thought they were being asked *how important exercise is*, not how important exercise is *to them, personally*. The researchers failed to explain properly to their respondents what the ladder was supposed to measure, and even devout couch potatoes are going to tell you that exercise is important if you ask them the general question.

The Faces Scale

Another interesting device is the faces scale shown in figure 12.3. It's a 7-point (or 5-point, or 9-point) scale with stylized faces that change from joy to gloom.

This technique was developed by Kunin in 1955 to measure job satisfaction

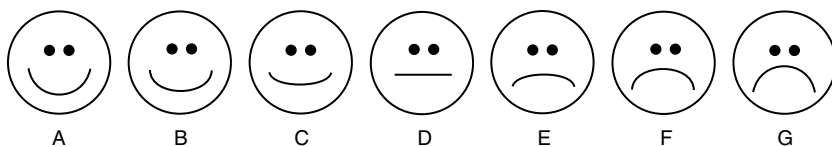


Figure 12.3. The faces scale.

SOURCE: F. M. Andrews and S. B. Withey, *Social Indicators of Well-Being: Americans' Perceptions of Life Quality*, Appendix A, p. 13. © 1976. Reprinted by permission of Springer.

and has been used widely for this ever since (Smith et al. 1969; Brief and Roberson 1989; Wanous et al. 1997). It's a really good device for capturing people's feelings about a wide variety of things—health care, personal safety, consumer items (brands of beer, titles of current movies), and so on. People are told: “Here are some faces expressing various feelings. Which face comes closest to how you feel about _____?” Try using this scale with names of well-known political figures or music artists just to get a feel for how interesting it is.

Physicians and psychologists use this scale as a prop when they ask patients to describe pain. It's particularly good when working with children, but it's effective with adults as well (Belter et al. 1988; Bieri et al. 1990; Harrison 1993).

There is some evidence that the meaning of the faces in figure 12.3 is nearly universal (Ekman et al. 1969), but Oliver Kortendick used the faces scale in his research on social networks in Papua New Guinea, and people did not respond well to the task. It seems that the face farthest to the right, which almost everyone in Europe, North America, and Latin America interprets as “unhappy” was interpreted in Papua New Guinea as “hostility” and “aggression”—two emotions that were simply not talked about openly in the village where Kortendick did his work (personal communication).

And Finally

There are thousands of published scales. Whatever you're interested in, the chances are good that someone has developed and tested a scale to measure it. Of course, scales are not automatically portable. A scale that measures stress among Barbadian women may not measure stress among Ghanaian men.

Still, it makes sense to seek out any published scales on variables you're studying. You may be able to adapt the scales to your needs, or you may get ideas for building and testing an alternative scale. Just because scales are not perfectly transportable across time and cultures doesn't mean those scales are useless to you. For a start on looking for scales that you can adapt, consult Miller and Salkind (2002) and Beere (1990). For more on developing scales, see Spector (1992), DeVellis (2003), Netemeyer et al. (2003), and Dunn-Rankin (2004). Some classics on scaling include Torgerson (1958), Nunnally and Bernstein (1994), and Coombs (1964).

13



Participant Observation

Participant observation fieldwork is the foundation of cultural anthropology. It involves getting close to people and making them feel comfortable enough with your presence so that you can observe and record information about their lives. If this sounds a bit crass, I mean it to come out that way. Only by confronting the truth about participant observation—that it involves deception and impression management—can we hope to conduct ourselves ethically in fieldwork. Much more about this later.

Participant observation is both a humanistic method and a scientific one. It produces the kind of experiential knowledge that lets you talk convincingly, from the gut, about what it feels like to plant a garden in the high Andes or dance all night in a street rave in Seattle.

It also produces effective, positivistic knowledge—the kind that can move the levers of the world if it gets into the right hands. Nancy Scheper-Hughes (1992), for example, developed a nomothetic theory, based on participant observation, that accounts for the tragedy of very high infant mortality in northeast Brazil and the direct involvement of mothers in their infants' deaths. Anyone who hopes to develop a program to lower the incidence of infant mortality in that part of the world will certainly have to read Scheper-Hughes's analysis.

And participant observation is used in product development and other direct applications research—that is, where the object from the start is to solve a human problem. Brigitte Jordan and her team of ethnographers at Xerox corporation determined the information flow and the hierarchy of interactions in the operations room of a major airline at a metropolitan airport (Jordan 1992b). And when credit-card readers were first installed on gasoline pumps in the early 1990s, consumers avoided using the technology. John Lowe and a team of participant observers figured out why (Solomon 1993).

Romancing the Methods

It used to be that the skills for doing fieldwork were mysterious and unteachable, something you just learned, out there in the field. In the 1930s, John Whiting and some of his fellow anthropology students at Yale University asked their professor, Leslie Spier, for a seminar on methods. “This was a subject to discuss casually at breakfast,” Whiting recalls Spier telling him, not something worthy of a seminar (Whiting 1982:156). Tell this story to seasoned anthropologists at a convention, and it’s a good bet they’ll come back with a story of their own just like it.

It’s fine for anthropologists to romanticize fieldwork—vulcanologists and oceanographers do it, too, by the way—particularly about fieldwork in places that take several days to get to, where the local language has no literary tradition, and where the chances are nontrivial of coming down with a serious illness. Research really is harder to do in some places than in others. But the fact is, anthropologists are more likely these days to study drug use among urban African Americans (Dei 2002), the daily life of the mentally retarded in a common residence (Angrosino 1997), the life of police in Los Angeles (Barker 1999), army platoons in Britain (Killworth 1997), consumer behavior (Sherry 1995), gay culture (Herdt 1992; Murray 1992), or life on the mean streets of big cities (Bourgeois 1995; Fleisher 1998) than they are to study isolated tribal or peasant peoples. It would take a real inventory to find out how much more likely, but in a recent collection of 17 self-reflective studies of anthropologists about their fieldwork (Hume and Mulcock 2004), just three cases deal with work in isolated communities. (For more on street ethnography, see Agar 1973, Weppner 1973, 1977, Fleisher 1995, Lambert et al. 1995, Connolly and Ennew 1996, Gigengack 2000, and Kane 2001.)

And while participant observation in small, isolated communities has some special characteristics, the techniques and skills that are required seem to me to be pretty much the same everywhere.

What *Is* Participant Observation?

Participant observation usually involves fieldwork, but not all fieldwork is participant observation. Goldberg et al. (1994) interviewed 206 prostitutes and collected saliva specimens (to test for HIV and for drug use) during 53 nights of fieldwork in Glasgow’s red light district. This was serious fieldwork, but hardly participant observation.

So much for what participant observation isn’t. Here’s what it is: Participant observation is one of those **strategic methods** I talked about in chapter

1—like experiments, surveys, or archival research. It puts you where the action is and lets you collect data . . . any kind of data you want, narratives or numbers. It has been used for generations by positivists and interpretivists alike.

A lot of the data collected by participant observers are qualitative: field notes taken about things you see and hear in natural settings; photographs of the content of people's houses; audio recordings of people telling folktales; videotapes of people making canoes, getting married, having an argument; transcriptions of taped, open-ended interviews, and so on.

But lots of data collected by participant observers are quantitative and are based on methods like direct observation, questionnaires, and pile sorts. Whether you consider yourself an interpretivist or a positivist, participant observation gets you in the door so you can collect life histories, attend rituals, and talk to people about sensitive topics.

Participant observation involves going out and staying out, learning a new language (or a new dialect of a language you already know), and experiencing the lives of the people you are studying as much as you can. Participant observation is about stalking culture in the wild—establishing rapport and learning to act so that people go about their business as usual when you show up. If you are a successful participant observer, you will know when to laugh at what people think is funny; and when people laugh at what you say, it will be because you *meant* it to be a joke.

Participant observation involves immersing yourself in a culture and learning to remove yourself every day from that immersion so you can intellectualize what you've seen and heard, put it into perspective, and write about it convincingly. When it's done right, participant observation turns fieldworkers into instruments of data collection and data analysis.

The implication is that *better* fieldworkers are *better* data collectors and *better* data analyzers. And the implication of *that* is that participant observation is not an attitude or an epistemological commitment or a way of life. It's a craft. As with all crafts, becoming a skilled artisan at participant observation takes practice.

Some Background and History

Bronislaw Malinowski (1884–1942) didn't invent participant observation, but he is widely credited with developing it as a serious method of social research. A British social anthropologist (born in Poland), Malinowski went out to study the people of the Trobriand Islands, in the Indian Ocean, just before World War I. At the time, the Trobriand Islands were a German posses-

sion, so when the war broke out, Malinowski was interned and could not return to England for three years.

He made the best of the situation, though. Here is Malinowski describing his methods:

Soon after I had established myself in Omarkana, Trobriand Islands, I began to take part, in a way, in the village life, to look forward to the important or festive events, to take personal interest in the gossip and the developments of the village occurrences; to wake up every morning to a new day, presenting itself to me more or less as it does to the natives. . . . As I went on my morning walk through the village, I could see intimate details of family life, of toilet, cooking, taking of meals; I could see the arrangements for the day's work, people starting on their errands, or groups of men and women busy at some manufacturing tasks.

Quarrels, jokes, family scenes, events usually trivial, sometimes dramatic but always significant, form the atmosphere of my daily life, as well as of theirs. It must be remembered that the natives saw me constantly every day, they ceased to be interested or alarmed, or made self-conscious by my presence, and I ceased to be a disturbing element in the tribal life which I was to study, altering it by my very approach, as always happens with a newcomer to every savage community. In fact, as they knew that I would thrust my nose into everything, even where a well-mannered native would not dream of intruding, they finished by regarding me as a part and parcel of their life, a necessary evil or nuisance, mitigated by donations of tobacco. (1961 [1922]:7–8)

Ignore the patronizing rhetoric about the “savage community” and “donations of tobacco.” (I’ve learned to live with this part of our history in anthropology. Knowing that all of us, in every age, look quaint, politically incorrect, or just plain hopeless to those who come later has made it easier.) Focus instead on the amazing, progressive (for that time) method that Malinowski advocated: Spend lots and lots of time in studying a culture, learn the language, hang out, do all the everyday things that everyone else does, become inconspicuous by sheer tenaciousness, and stay aware of what’s really going on. Apart from the colonialist rhetoric, Malinowski’s discussion of participant observation is as resonant today as it was more than 80 years ago.

By the time Malinowski went to the Trobriands, *Notes and Queries on Anthropology*—the fieldwork manual produced by the Royal Anthropological Institute of Great Britain and Ireland—was in its fourth edition. The first edition came out in 1874 and the last edition (the sixth) was reprinted five times until 1971.

Thirty-five years later, that final edition of *Notes and Queries* is still must reading for anyone interested in learning about anthropological field methods. Once again, ignore the fragments of paternalistic colonialism—“a sporting

rifle and a shotgun are . . . of great assistance in many districts where the natives may welcome extra meat in the shape of game killed by their visitor” (Royal Anthropological Institute 1951:29)—and *Notes and Queries* is full of useful, late-model advice about how to conduct a census, how to handle photographic negatives in the field, and what questions to ask about sexual orientation, infanticide, food production, warfare, art. . . . The book is just a treasure.

We make the most consistent use of participant observation in anthropology, but the method has very, very deep roots in sociology. Beatrice Webb was doing participant observation—complete with note taking and informant interviewing—in the 1880s and she wrote trenchantly about the method in her 1926 memoir (Webb 1926). Just about then, the long tradition in sociology of urban ethnography—the “Chicago School”—began at the University of Chicago under the direction of Robert Park and Ernest Burgess (see Park et al. 1925). One of Park’s students was his son-in-law, Robert Redfield, the anthropologist who pioneered community studies in Mexico.

Just back from lengthy fieldwork with Aborigine peoples in Australia, another young anthropologist, William Lloyd Warner, was also influenced by Park. Warner launched one of the most famous American community-study projects of all time, the Yankee City series (Warner and Hunt 1941; Warner 1963). (Yankee City was the pseudonym for Newburyport, Massachusetts.) In 1929, sociologists Robert and Helen Lynd published the first of many ethnographies about Middletown. (Middletown was the pseudonym for Muncie, Indiana.)

Some of the classic ethnographies that came out of the early Chicago School include Harvey Zorbaugh’s *The Gold Coast and the Slum* (1929) and Clifford Shaw’s *The Jack-Roller* (1930). In *The Jack-Roller*, a 22 year old named Stanley talks about what it was like to grow up as a delinquent in early 20th-century Chicago. It still makes great reading.

Becker et al.’s *Boys in White* (1961)—about the student culture of medical school in the 1950s—should be required reading, even today, for anyone trying to understand the culture of medicine in the United States. The ethnography tradition in sociology continues in the pages of the *Journal of Contemporary Ethnography*, which began in 1972 under the title *Urban Life and Culture*. (See Lofland [1983] and Bulmer [1984] for more on the history of the Chicago School of urban ethnography.)

Participant observation today is everywhere—in political science, management, education, nursing, criminology, social psychology—and one of the terrific results of all this is a growing body of literature about participant observation itself. There are highly focused studies, full of practical advice, and there are poignant discussions of the overall *experience* of fieldwork. For large

doses of both, see Wolcott (1995), Agar (1996), and C. D. Smith and Kornblum (1996), Handwerker (2001), and Dewalt and Dewalt (2002). There's still plenty of mystery and romance in participant observation, but you don't have to go out unprepared.

Fieldwork Roles

Fieldwork can involve three very different roles: (1) **complete participant**, (2) **participant observer**, and (3) **complete observer**. The first role involves deception—becoming a member of a group without letting on that you're there to do research. The third role involves following people around and recording their behavior with little if any interaction. This is part of direct observation, which we'll take up in the next chapter.

By far, most ethnographic research is based on the second role, that of the participant observer. Participant observers can be insiders who observe and record some aspects of life around them (in which case, they're **observing participants**); or they can be outsiders who participate in some aspects of life around them and record what they can (in which case, they're **participating observers**).

In 1965, I went to sea with a group of Greek sponge fishermen in the Mediterranean. I lived in close quarters with them, ate the same awful food as they did, and generally participated in their life—as an outsider. I didn't dive for sponges, but I spent most of my waking hours studying the behavior and the conversation of the men who did. The divers were curious about what I was writing in my notebooks, but they went about their business and just let me take notes, time their dives, and shoot movies (Bernard 1987). I was a **participating observer**.

Similarly, when I went to sea in 1972 and 1973 with oceanographic research vessels, I was part of the scientific crew, there to watch how oceanographic scientists, technicians, and mariners interacted and how this interaction affected the process of gathering oceanographic data. There, too, I was a participating observer (Bernard and Killworth 1973).

Circumstances can sometimes overtake the role of mere participating observer. In 1979, El Salvador was in civil war. Thousands fled to Honduras where they were sheltered in refugee camps near the border. Phillipe Bourgois went to one of those camps to initiate what he hoped would be his doctoral research in anthropology. Some refugees there offered to show him their home villages and Bourgois crossed with them, illegally, into El Salvador for what he thought would be a 48-hour visit. Instead, Bourgois was trapped, along with about a thousand peasants, for 2 weeks, as the Salvadoran military bombed,

shelled, and strafed a 40-square-kilometer area in search of rebels (Bourgois 1990).

Mark Fleisher (1989) studied the culture of guards at a federal penitentiary in California, but as an observing participant, an insider. Researchers at the U.S. Federal Bureau of Prisons asked Fleisher to do an ethnographic study of job pressures on guards—called correctional officers, or COs in the jargon of the profession—in a maximum-security federal penitentiary. It costs a lot to train a CO, and there was an unacceptably high rate of them leaving the job after a year or two. Could Fleisher look into the problem?

Fleisher said he'd be glad to do the research and asked when he could start "walking the mainline"—that is, accompanying the COs on their rounds through the prison. He was told that he'd be given an office at the prison and that the guards would come to his office to be interviewed.

Fleisher said he was sorry, but he was an anthropologist, he was doing participant observation, and he'd have to have the run of the prison. Sorry, they said back, only sworn correctional officers can walk the prison halls. So, swear me in, said Fleisher, and off he went to training camp for 6 weeks to become a sworn federal correctional officer. *Then* he began his yearlong study of the U.S. Penitentiary at Lompoc, California. In other words, he became an observing participant in the culture he was studying. Fleisher never hid what he was doing. When he went to USP-Lompoc, he told everyone that he was an anthropologist doing a study of prison life.

Barbara Marriott (1991) studied how the wives of U.S. Navy male officers contributed to their husbands' careers. Marriott was herself the wife of a retired captain. She was able to bring the empathy of 30 years' full participation to her study. She, too, took the role of observing participant and, like Fleisher, she told her informants exactly what she was doing.

Holly Williams (1995) spent 14 years as a nurse, ministering to the needs of children who had cancer. When Williams did her doctoral dissertation, on how the parents of those young patients coped with the trauma, she started as a credible insider, as someone whom the parents could trust with their worst fears and their hopes against all hope. Williams was a complete participant who became an observing participant by telling the people whom she was studying exactly what she was up to and enlisting their help with the research.

Going Native

Some fieldworkers start out as participating observers and find that they are drawn completely into their informants' lives. In 1975, Kenneth Good went to study the Yanomami in the Venezuelan Amazon. He planned on living with the Yanomami for 15 months, but he stayed on for nearly 13 years. "To my

great surprise,” says Good, “I had found among them a way of life that, while dangerous and harsh, was also filled with camaraderie, compassion, and a thousand daily lessons in communal harmony” (Good 1991:ix). Good learned the language and became a nomadic hunter and gatherer. He was adopted into a lineage and given a wife. (Good and his wife, Yárima, tried living in the United States, but after a few years, Yárima returned to the Yanomami.)

Marlene Dobkin de Rios did fieldwork in Peru and married the son of a Peruvian folk healer, whose practice she studied (Dobkin de Rios 1981). And Jean Gearing (1995) is another anthropologist who married her closest informant on the island of St. Vincent.

Does going native mean loss of objectivity? Perhaps, but not necessarily. In the industrialized countries of the West—the United States, Canada, Germany, Australia, Germany, England, France, etc.—we *expect* immigrants to go native. We expect them to become fluent in the local language, to make sure that their children become fully acculturated, to participate in the economy and politics of the nation, and so on.

Some fully assimilated immigrants to those countries become anthropologists and no one questions whether their immigrant background produces a lack of objectivity. Since total objectivity is, by definition, a myth, I’d worry more about producing credible data and strong analysis and less about whether going native is good or bad.

How Much Time Does It Take?

Anthropological field research traditionally takes a year or more because it takes that long to get a feel for the full round of people’s lives. It can take that long just to settle in, learn a new language, gain rapport, and be in a position to ask good questions and to get good answers.

A lot of participant observation studies, however, are done in a matter of weeks or a few months. Yu (1995) spent 4 months as a participant observer in a family-run Chinese restaurant, looking at differences in the conceptions that Chinese and non-Chinese employees had about things like good service, adequate compensation, and the role of management.

At the extreme low end, it is possible to do useful participant observation in just a few days. Assuming that you’ve wasted as much time in laundromats as I did when I was a student, you could conduct a reasonable participant observation study of one such place in a week. You’d begin by bringing in a load of wash and paying careful attention to what’s going on around you.

After two or three nights of observation, you’d be ready to tell other patrons that you were conducting research and that you’d appreciate their letting you

interview them. The reason you could do this is because you already speak the native language and have already picked up the nuances of etiquette from previous experience. Participant observation would help you intellectualize what you already know.

In general, though, participant observation is not for the impatient. Gerald Berreman studied life in Sirkanda, a Pahari-speaking village in north India. Berreman's interpreter-assistant, Sharma, was a Hindu Brahmin who neither ate meat nor drank alcohol. As a result, villagers did neither around Berreman or his assistant. Three months into the research, Sharma fell ill and Berreman hired Mohammed, a young Muslim schoolteacher to fill in.

When the villagers found out that Mohammed ate meat and drank alcohol, things broke wide open and Berreman found out that there were frequent inter-caste meat and liquor parties. When villagers found out that the occasional drink of locally made liquor was served at Berreman's house "access to information of many kinds increased proportionately" (Berreman 1962:10). Even then, it still took Berreman 6 months in Sirkanda before people felt comfortable performing animal sacrifices when he was around (*ibid.*:20).

And don't think that long term is only for foreign fieldwork. It took Daniel Wolf 3 years just to get into the Rebels, a brotherhood of outlaw bikers, and another couple of years riding with them before he had the data for his doctoral dissertation (Wolf 1991).

The amount of time you spend in the field can make a big difference in what you learn. Raoul Naroll (1962) found that anthropologists who stayed in the field for at least a year were more likely to report on sensitive issues like witchcraft, sexuality, political feuds, etc. Back in chapter 3, I mentioned David Price's study of water theft among farmers in Egypt's Fayoum Oasis. You might have wondered then how in the world he was able to do that study. Each farmer had a water allotment—a certain day each week and a certain amount of time during which water could flow to his fields. Price lived with these farmers for 8 months before they began telling him privately that they occasionally diverted water to their own fields from those of others (1995:106). Ethnographers who have done very long-term participant observation—that is, a series of studies over decades—find that they eventually get data about social change that is simply not possible to get in any other way (Kemper and Royce 2002).

My wife Carole and I spent May 2000 on Kalymnos, the Greek island where I did my doctoral fieldwork in 1964–1965. We've been visiting that island steadily for 40 years, but something qualitatively different happened in 2000. I couldn't quite put my finger on it, but by the end of the month I realized that people were talking to me about grandchildren. The ones who had grandchildren were chiding me—very good-naturedly, but chiding nonethe-

less—for not having any grandchildren yet. The ones who didn't have grandchildren were in commiseration mode. They wanted someone with whom to share their annoyance that “Kids these days are in no hurry to make families” and that “All kids want today . . . especially girls . . . is to have careers.”

This launched lengthy conversations about how “everything had changed” since we had been our children's ages and about how life in Greece was getting to be more and more like Europe (which is what many Greeks call Germany, France, and the rest of the fully industrialized nations of the European Union), and even like the United States. I suppose there were other ways I could have gotten people into give-and-take conversations about culture change, gender roles, globalization, modernization, and other big topics, but the grandchildren deficit was a terrific opener in 2000. And the whole conversation would have been a nonstarter had I been 30 instead of 60 years old. It wasn't just age, by the way; it was the result of the rapport that comes with having common history with people.

Here's history. In 1964, Carole and I brought our then 2-month-old daughter with us. Some of the same people who joked with me in 2000 about not having grandchildren had said to me in 1964: “Don't worry, next time you'll have a son.” I recall having been really, really annoyed at the time, but writing it down as data. A couple of years later, I sent friends on Kalymnos the announcement of our second child—another girl. I got back kidding remarks like “Congratulations! Keep on trying. . . . Still plenty of time to have a boy!” That was data, too. And when I told people that Carole and I had decided to stop at two, some of them offered mock condolences: “Oh, now you're really in for it! You'll have to get dowries for *two* girls without any sons to help.” Now *that's* data!

Skip to 2004, when our daughter, son-in-law, and new granddaughter Zoë came to Kalymnos for Zoë's first birthday. There is a saying in Greek that “the child of your child is two times your child.” You can imagine all the conversations, late into the night, about that. More data.

Bottom line: You can do highly focused participant observation research in your own language, to answer specific questions about your own culture, in a short time. How do middle-class, second-generation Mexican American women make decisions on which of several brands of pinto beans to select when they go grocery shopping? If you are a middle-class Mexican American woman, you can probably find the answer to that question, using participant observation, in a few weeks, because you have a wealth of personal experience to draw on.

But if you're starting out fresh, and not as a member of the culture you're studying, count on taking 3 months or more, under the best conditions, to be

accepted as a participant observer—that is, as someone who has learned enough to learn. And count on taking a lifetime to learn some things.

Rapid Assessment

Applied ethnographic research is often done in just a few weeks. Applied researchers just don't have the luxury of doing long-term participant observation fieldwork and may use **rapid assessment** procedures, especially **participatory rapid assessment**, or **PRA**. PRA (of agricultural or medical practices, for example) may include participant observation.

Rapid assessment means going in and getting on with the job of collecting data without spending months developing rapport. This means going into a field situation armed with a list of questions that you want to answer and perhaps a checklist of data that you need to collect.

Chambers (1991) advocates **participatory mapping**. He asks people to draw maps of villages and to locate key places on the maps. In **participatory transects**, he borrows from wildlife biology and systematically walks through an area, with key informants, observing and asking for explanations of everything he sees along the transect. He engages people in group discussions of key events in a village's history and asks them to identify clusters of households according to wealth. In other words, as an applied anthropologist, Chambers is called on to do rapid assessment of rural village needs, and he takes the people fully into his confidence as research partners. This method is just as effective in organizations as in small villages.

Applied medical anthropologists also use rapid assessment methods. The focused ethnographic study method, or FES, was developed by Sandy Gove (a physician) and Gretel Pelto (an anthropologist) for the World Health Organization to study acute respiratory illness (ARI) in children. The FES manual gives detailed instructions to fieldworkers for running a rapid ethnographic study of ARI in a community (WHO 1993; Gove and Pelto 1994).

Many ARI episodes turn out to be what physicians call pneumonia, but that is not necessarily what mothers call the illness. Researchers ask mothers to talk about recent ARI events in their households. Mothers also free list the symptoms, causes, and cures for ARI and do pile sorts of illnesses to reveal the folk taxonomy of illness and where ARI fits into that taxonomy. There is also a matching exercise, in which mothers pair locally defined symptoms (fever, sore throat, headache . . .) with locally defined causes (bad water, evil eye, germs . . .), cures (give rice water, rub the belly, take child to the doctor . . .), and illnesses.

The FES method also uses vignettes, or scenarios, much like those devel-

oped by Peter Rossi for the factorial survey (see chapter 10). Mothers are presented with cases in which variables are changed systematically (“Your child wakes up with [mild] [strong] fever. He complains that he has [a headache] [stomach ache],” and so on) and are asked to talk about how they would handle the case.

All this evidence—the free narratives, the pile sorts, the vignettes, etc.—is used in understanding the emic part of ARI, the local explanatory model for the illness.

Researchers also identify etic factors that make it easy or hard for mothers to get medical care for children who have pneumonia. These are things like the distance to a clinic, the availability of transportation, the number of young children at home, the availability to mothers of people with whom they can leave their children for a while, and so on. (For an example of the FES in use, see Hudelson 1994.)

The key to high-quality, quick ethnography, according to Handwerker (2001), is to go into a study with a clear question and to limit your study to five focus variables. If the research is exploratory, you just have to make a reasonable guess as to what variables might be important and hope for the best. Most rapid assessment studies, however, are applied research, which usually means that you can take advantage of earlier, long-term studies to narrow your focus.

For example, Edwins Laban Moogi Gwako (1997) spent over a year testing the effects of eight independent variables on Maragoli women’s agricultural productivity in western Kenya. At the end of his doctoral research, he found that just two variables—women’s land tenure security and the total value of their household wealth—accounted for 46% of the variance in productivity of plots worked by women. None of the other variables—household size, a woman’s age, whether a woman’s husband lived at home, and so on—had any effect on the dependent variable.

If you were doing a rapid assessment of women’s agricultural productivity elsewhere in east Africa, you would take advantage of Laban Moogi Gwako’s work and limit the variables you tested to perhaps four or five—the two that he found were important and perhaps two or three others. You can study this same problem for a lifetime, and the more time you spend, the more you’ll understand the subtleties and complexities of the problem. But the point here is that if you have a clear question and a few, clearly defined variables, you can produce quality work in a lot less time than you might imagine. For more on rapid ethnographic assessment, see Bentley et al. (1988), Scrimshaw and Hurtado (1987), and Scrimshaw and Gleason (1992). See Baker (1996a, 1996b) for a PRA study of homeless children in Kathmandu.

Validity—Again

There are at least five reasons for insisting on participant observation in the conduct of scientific research about cultural groups.

1. Participant observation opens things up and makes it possible to collect all kinds of data. Participant observation fieldworkers have witnessed births, interviewed violent men in maximum-security prisons, stood in fields noting the behavior of farmers, trekked with hunters through the Amazon forest in search of game, and pored over records of marriages, births, and deaths in village churches and mosques around the world.

It is impossible to imagine a complete stranger walking into a birthing room and being welcomed to watch and record the event or being allowed to examine any community's vital records at whim. It is impossible, in fact, to imagine a stranger doing *any* of the things I just mentioned or the thousands of other intrusive acts of data collection that fieldworkers engage in all the time. What makes it all possible is participant observation.

2. Participant observation reduces the problem of **reactivity**—of people changing their behavior when they know that they are being studied. As you become less and less of a curiosity, people take less and less interest in your comings and goings. They go about their business and let you do such bizarre things as conduct interviews, administer questionnaires, and even walk around with a stopwatch, clipboard, and camera.

Phillipe Bourgois (1995) spent 4 years living in El Barrio (the local name for Spanish Harlem) in New York City. It took him a while, but eventually he was able to keep his tape recorder running for interviews about dealing crack cocaine and even when groups of men bragged about their involvement in gang rapes.

Margaret Graham (2003) weighed every gram of every food prepared for 75 people eating over 600 meals in 15 households in the Peruvian Andes. This was completely alien to her informants, but after 5 months of intimate participant observation, those 15 families allowed her to visit them several times, with an assistant and a food scale.

In other words: Presence builds trust. Trust lowers reactivity. Lower reactivity means higher validity of data. Nothing is guaranteed in fieldwork, though. Graham's informants gave her permission to come weigh their food, but the act of doing so turned out to be more alienating than either she or her informants had anticipated. By local rules of hospitality, people had to invite Graham to eat with them during the three visits she made to their homes—but

Graham couldn't accept any food, lest doing so bias her study of the nutritional intake of her informants. Graham discussed the awkward situation openly with her informants, and made spot checks of some families a few days after each weighing episode to make sure that people were eating the same kinds and portions of food as Graham had witnessed (Graham 2003:154).

And when Margaret LeCompte told children at a school that she was writing a book about them, they started acting out in "ways they felt would make good copy" by mimicking characters on popular TV programs (LeCompte et al. 1993).

3. Participant observation helps you ask sensible questions, in the native language. Have you ever gotten a questionnaire in the mail and said to yourself: "What a dumb set of questions"? If a social scientist who is a member of your own culture can make up what you consider to be dumb questions, imagine the risk *you* take in making up a questionnaire in a culture very different from your own! Remember, it's just as important to ask sensible questions in a face-to-face interview as it is on a survey instrument.
4. Participant observation gives you an intuitive understanding of what's going on in a culture and allows you to speak with confidence about the meaning of data. Participant observation lets you make strong statements about cultural facts that you've collected. It extends both the internal and the external validity of what you learn from interviewing and watching people. In short, participant observation helps you understand the *meaning* of your observations. Here's a classic example.

In 1957, N. K. Sarkar and S. J. Tambiah published a study, based on questionnaire data, about economic and social disintegration in a Sri Lankan village. They concluded that about two-thirds of the villagers were landless. The British anthropologist, Edmund Leach, did not accept that finding (Leach 1967). He had done participant observation fieldwork in the area, and knew that the villagers practiced patrilocal residence after marriage. By local custom, a young man might receive *use* of some of his father's land even though legal ownership might not pass to the son until the father's death.

In assessing land ownership, Sarkar and Tambiah asked whether a "household" had any land, and if so, how much. They defined an independent household as a unit that cooked rice in its own pot. Unfortunately, all married women in the village had their own rice pots. So, Sarkar and Tambiah wound up estimating the number of independent households as very high and the number of those households that owned land as very low. Based on these data, they concluded that there was gross inequality in land ownership and that this characterized a "disintegrating village" (the title of their book).

Don't conclude from Leach's critique that questionnaires are "bad," while

participant observation is “good.” I can’t say often enough that participant observation makes it possible to collect quantitative survey data or qualitative interview data from some sample of a population. Qualitative and quantitative data inform each other and produce insight and understanding in a way that cannot be duplicated by either approach alone. Whatever data collection methods you choose, participant observation maximizes your chances for making valid statements.

5. Many research problems simply cannot be addressed adequately by anything except participant observation. If you want to understand how a local court works, you can’t very well disguise yourself and sit in the courtroom unnoticed. The judge would soon spot you as a stranger, and, after a few days, you would have to explain yourself. It is better to explain yourself at the beginning and get permission to act as a participant observer. In this case, your participation consists of acting like any other local person who might sit in on the court’s proceedings. After a few days, or weeks, you would have a pretty good idea of how the court worked: what kinds of crimes are adjudicated, what kinds of penalties are meted out, and so forth. You might develop some specific hypotheses from your qualitative notes—hypotheses regarding covariations between severity of punishment and independent variables other than severity of crime. Then you could test those hypotheses on a sample of courts.

Think this is unrealistic? Try going down to your local traffic court and see whether defendants’ dress or manner of speech predict variations in fines for the same infraction. The point is, getting a general understanding of how any social institution or organization works—the local justice system, a hospital, a ship, or an entire community—is best achieved through participant observation.

Entering the Field

Perhaps the most difficult part of actually doing participant observation fieldwork is making an entry. There are five rules to follow.

1. There is no reason to select a site that is difficult to enter when equally good sites are available that are easy to enter (see chapter 3). In many cases, you *will* have a choice—among equally good villages in a region, or among school districts, hospitals, or cell blocks. When you have a choice, take the field site that promises to provide easiest access to data.
2. Go into the field with plenty of written documentation about yourself and your project. You’ll need formal letters of introduction—at a minimum, from your university, or from your client if you are doing applied work on a contract. Let-

ters from universities should spell out your affiliation, who is funding you, and how long you will be at the field site.

Be sure that those letters are in the language spoken where you will be working, and that they are signed by the highest academic authorities possible.

Letters of introduction should not go into detail about your research. Keep a separate document handy in which you describe your proposed work, and present it to gatekeepers who ask for it, along with your letters of introduction.

Of course, if you study an outlaw biker gang, like Daniel Wolf did, forget about letters of introduction (Wolf 1991).

3. Don't try to wing it, unless you absolutely have to. There is nothing to be said for "getting in on your own." Use personal contacts to help you make your entry into a field site.

When I went to Kalymnos, Greece, in 1964, I carried with me a list of people to look up. I collected the list from people in the Greek American community of Tarpon Springs, Florida, who had relatives on Kalymnos. When I went to Washington, D.C., to study how decision makers in the bureaucracy used (or didn't use) scientific information, I had letters of introduction from colleagues at Scripps Institution of Oceanography (where I was working at the time).

If you are studying any hierarchically organized community (hospitals, police departments, universities, school systems, etc.), it is usually best to start at the top and work down. Find out the names of the people who are the gatekeepers and see them first. Assure them that you will maintain strict confidentiality and that no one in your study will be personally identifiable.

In some cases, though, starting at the top can backfire. If there are warring factions in a community or organization, and if you gain entry to the group at the top of *one* of those factions, you will be asked to side with that faction.

Another danger is that top administrators of institutions may try to enlist you as a kind of spy. They may offer to facilitate your work if you will report back to them on what you find out about specific individuals. This is absolutely off limits in research. If that's the price of doing a study, you're better off choosing another institution. In the 2 years I spent doing research on communication structures in federal prisons, no one ever asked me to report on the activities of specific inmates. But other researchers have reported experiencing this kind of pressure, so it's worth keeping in mind.

4. Think through in advance what you will say when ordinary people (not just gatekeepers) ask you: What are you doing here? Who sent you? Who's funding you? What good is your research and who will it benefit? Why do you want to learn

about people here? How long will you be here? How do I know you aren't a spy for _____? (where the blank is filled in by whoever people are afraid of).

The rules for presentation of self are simple: Be honest, be brief, and be absolutely consistent. In participant observation, if you try to play any role other than yourself, you'll just get worn out (Jones 1973).

But understand that not everyone will be thrilled about your role as a researcher. Terry Williams studied cocaine use in after-hours clubs in New York. It was "gay night" in one bar he went to. Williams started a conversation with a man whose sleeves were fully rolled, exposing tattoos on both arms. The man offered to buy Williams a drink. Was this Williams's first time at the bar? Williams said he'd been there before, that he was a researcher, and that he just wanted to talk. The man turned to his friends and exploded: "Hey, get a load of this one. He wants to do research on us. You scum bag! What do we look like, pal? Fucking guinea pigs?" (Williams 1996:30).

After that experience, Williams became, as he said, "more selective" in whom he told about his real purpose in those after-hours clubs.

5. Spend time getting to know the physical and social layout of your field site. It doesn't matter if you're working in a rural village, an urban enclave, or a hospital. Walk it and write notes about how it *feels* to you. Is it crowded? Do the buildings or furniture seem old or poorly kept? Are there any distinctive odors?

You'd be surprised how much information comes from asking people about little things like these. I can still smell the distinctive blend of diesel fuel and taco sauce that's characteristic of so many bus depots in rural Mexico. Asking people about those smells opened up long conversations about what it's like for poor people, who don't own cars, to travel in Mexico and all the family and business reasons they have for traveling. If something in your environment makes a strong sensory impression, write it down.

A really good early activity in any participant observation project is to make maps and charts—kinship charts of families, chain-of-command charts in organizations, maps of offices or villages or whatever physical space you're studying, charts of who sits where at meetings, and so on.

For making maps, take a GPS (global positioning system) device to the field with you. They are small, easy to use, and relatively inexpensive. GPS devices that are accurate to within 10 meters are available for under \$200 (see appendix F for more). What a GPS does is track your path via satellite, so that if you can walk the perimeter of an area, you can map it and mark its longitude and latitude accurately. Eri Sugita (2004) studied the relation between the

washing of hands by the mothers of young children and the rate of diarrheal disease among those children in Bugobero, Uganda. Sugita used a GPS device to map the position of every well and every spring in Bugobero. Then she walked to each of the water sources from each of the 51 households in her study and, wearing a pedometer, measured the travel distance to the nearest source of clean water. (You can also make maps using multidimensional scaling. See chapter 21. For more on pedometers, see Tudor-Locke et al. 2004.)

Another good thing to do is to take a census of the group you're studying as soon as you can. When she began her fieldwork on the demography and fertility in a Mexican village, Julia Pauli (2000) did a complete census of 165 households. She recorded the names of all the people who were considered to be members of the household, whether they were living there or not (a lot of folks were away, working as migrant laborers). She recorded their sex, age, religion, level of education, marital status, occupation, place of birth, and where each person was living right then. Then, for each of the 225 women who had given birth at least once, she recorded the name, sex, birth date, education, current occupation, marital status, and current residence of each child.

Pauli gave each person in a household their own, unique identification number and she gave each child of each woman in a household an I.D. number—whether the child was living at home, away working, or married and living in a separate household in the village. In the course of her census, she would eventually run into those married children living in other households. But since each person kept his or her unique I.D. number, Pauli was able to link all those born in the village to their natal homes. In other words, Pauli used the data from her straightforward demographic survey to build a kinship network of the village.

A census of a village or a hospital gives you the opportunity to walk around a community and to talk with most of its members at least once. It lets you be seen by others and it gives you an opportunity to answer questions, as well as to ask them. It allows you to get information that official censuses don't retrieve. And it can be a way to gain rapport in a community. But it can also backfire if people are afraid you might be a spy. Michael Agar reports that he was branded as a Pakistani spy when he went to India, so his village census was useless (1980b).

The Skills of a Participant Observer

To a certain extent, participant observation must be learned in the field. The strength of participant observation is that you, as a researcher, become the instrument for data collection and analysis through your own experience. Con-

sequently, you have to experience participant observation to get good at it. Nevertheless, there are a number of skills that you can develop before you go into the field.

Learning the Language

Unless you are a full participant in the culture you're studying, being a participant observer makes you a freak. Here's how anthropologists looked to Vine Deloria (1969:78), a Sioux writer:

Anthropologists can readily be identified on the reservations. Go into any crowd of people. Pick out a tall gaunt white man wearing Bermuda shorts, a World War II Army Air Force flying jacket, an Australian bush hat, tennis shoes, and packing a large knapsack incorrectly strapped on his back. He will invariably have a thin, sexy wife with stringy hair, an I.Q. of 191, and a vocabulary in which even the prepositions have eleven syllables. . . . This creature is an anthropologist.

Now, nearly four decades later, it's more likely to be the anthropologist's husband who jabbars in 11-syllable words, but the point is still the same. The most important thing you can do to stop being a freak is to speak the language of the people you're studying—and speak it well. Franz Boas was adamant about this. “Nobody,” he said, “would expect authoritative accounts of the civilization of China or Japan from a man who does not speak the languages readily, and who has not mastered their literatures” (1911:56). And yet, “the best kept secret of anthropology,” says Robbins Burling, “is the linguistic incompetence of ethnological fieldworkers” (2000 [1984]:v; and see Owusu [1978]; Werner [1994]; Borchgrevink [2003]).

That secret is actually not so well kept. In 1933, Paul Radin, one of Franz Boas's students, complained that Margaret Mead's work on Samoa was superficial because she wasn't fluent in Samoan (Radin 1966 [1933]:179). Sixty-six years later, Derek Freeman (1999) showed that Mead was probably duped by at least some of her adolescent informants about the extent of their sexual experience because she didn't know the local language.

In fact, Mead talked quite explicitly about her use of interpreters. It was not necessary, said Mead, for fieldworkers to become what she called “virtuosos” in a native language. It was enough to “use” a native language, as she put it, without actually speaking it fluently:

If one knows how to exclaim “how beautiful!” of an offering, “how fat!” of a baby, “how big!” of a just shot pig; if one can say “my foot's asleep” or “my back itches” as one sits in a closely packed native group with whom one is as yet unable to hold a sustained conversation; if one can ask the simple questions: “Is

that your child?" "Is your father living?" "Are the mosquitoes biting you?" or even utter culturally appropriate squeals and monosyllables which accompany fright at a scorpion, or startle at a loud noise, it is easy to establish rapport with people who depend upon affective contact for reassurance. (Mead 1939:198)

Robert Lowie would have none of it. A people's ethos, he said, is never directly observed. "It can be inferred only from their self-revelations," and this, indeed, requires the dreaded virtuosity that Mead had dismissed (Lowie 1940:84ff). The "horse-and-buggy ethnographers," said Lowie, in a direct response to Mead in the *American Anthropologist*, accepted virtuosity—that is, a thorough knowledge of the language in which one does fieldwork—on principle. "The new, stream-lined ethnographers," he taunted, rejected this as superfluous (ibid.:87). Lowie was careful to say that a thorough knowledge of a field language did not mean native proficiency. And, of course, Mead understood the benefits of being proficient in a field language. But she also understood that a lot of ethnography gets done through interpreters or through contact languages, like French, English, and pidgins . . . the not-so-well kept secret in anthropology.

Still . . . according to Brislin et al. (1973:70), Samoa is one of those cultures where "it is considered acceptable to deceive and to 'put on' outsiders. Interviewers are likely to hear ridiculous answers, not given in a spirit of hostility but rather sport." Brislin et al. call this the **sucker bias**, and warn fieldworkers to watch out for it. Presumably, knowing the local language fluently is one way to become alert to and avoid this problem.

And remember Raoul Naroll's finding that anthropologists who spent at least a year in the field were more likely to report on witchcraft? Well, he also found that anthropologists who spoke the local language were more likely to report data about witchcraft than were those who didn't. Fluency in the local language doesn't just improve your rapport; it increases the probability that people will tell you about sensitive things, like witchcraft, and that even if people try to put one over on you, you'll know about it (Naroll 1962:89–90).

When it comes to doing effective participant observation, learning a new jargon in your own language is just as important as learning a foreign language. Peggy Sullivan and Kirk Elifson studied the Free Holiness church, a rural group of Pentecostals whose rituals include the handling of poisonous snakes (rattles, cottonmouths, copperheads, and water moccasins). They had to learn an entirely new vocabulary:

Terms and expressions like "annointment," "tongues," "shouting," and "carried away in the Lord" began having meaning for us. We learned informally and often contextually through conversation and by listening to sermons and testimonials. The development of our understanding of the new language was gradual and

probably was at its greatest depth when we were most submerged in the church and its culture. . . . We simplified our language style and eliminated our use of profanity. We realized, for example, that one badly placed “damn” could destroy trust that we had built up over months of hard work. (Sullivan and Elifson 1996:36)

HOW TO LEARN A NEW LANGUAGE

In my experience, the way to learn a new language is to learn a few words and to say them brilliantly. Yes, study the grammar and vocabulary, but the key to learning a new language is saying things right, even just a handful of things. This means capturing not just the pronunciation of words, but also the intonation, the use of your hands, and other nonverbal cues that show you are really, really serious about the language and are trying to look and sound as much like a native as possible.

When you say the equivalent of “hey, hiya doin’” in any language—Zulu or French or Arabic—with just the right intonation, people will think you know more than you do. They’ll come right back at you with a flurry of words, and you’ll be lost. Fine. Tell them to slow down—again, in that great accent you’re cultivating.

Consider the alternative: You announce to people, with the first, badly accented words out of your mouth, that you know next to nothing about the language and that they should therefore speak to you with that in mind. When you talk to someone who is not a native speaker of your language, you make an automatic assessment of how large their vocabulary is and how fluent they are. You adjust both the speed of your speech and your vocabulary to ensure comprehension. That’s what Zulu and Arabic speakers will do with you, too. The trick is to act in a way that gets people into pushing your limits of fluency and into teaching you cultural insider words and phrases.

The real key to learning a language is to acquire vocabulary. People will usually figure out what you want to say if you butcher the grammar a bit, but they need nouns and verbs to even begin the effort. This requires studying lists of words every day and using as many new words every day as you can engineer into a conversation. Try to stick at least one conspicuously idiomatic word or phrase into your conversation every day. That will not only nail down some insider vocabulary, it will stimulate everyone around you to give you more of the same.

A good fraction of any culture is in the idioms and especially in the metaphors (more about metaphors in the section on schemata in chapter 17). To understand how powerful this can be, imagine you are hired to tutor a student from Nepal who wants to learn English. You point to some clouds and say

“clouds” and she responds by saying “clouds.” You say “very good” and she says “no brainer.” You can certainly pick up the learning pace after that kind of response.

As you articulate more and more insider phrases like a native, people will increase the rate at which they teach you by raising the level of their discourse with you. They may even compete to teach you the subtleties of their language and culture. When I was learning Greek in 1960 on a Greek merchant ship, the sailors took delight in seeing to it that my vocabulary of obscenities was up to their standards and that my usage of that vocabulary was suitably robust.

To prepare for my doctoral fieldwork in 1964–1965, I studied Greek at the University of Illinois. By the end of 1965, after a year on the island of Kalymnos, my accent, mannerisms, and vocabulary were more Kalymnian than Athenian. When I went to teach at the University of Athens in 1969, my colleagues there were delighted that I wanted to teach in Greek, but they were conflicted about my accent. How to reconcile the fact that an educated foreigner spoke reasonably fluent Greek with what they took to be a rural, working-class accent? It didn’t compute, but they were very forgiving. After all, I *was* a foreigner, and the fact that I was making an attempt to speak the local language counted for a lot.

So, if you are going off to do fieldwork in a foreign language, try to find an intensive summer course in the country where that language is spoken. Not only will you learn the language (and the local dialect of that language), you’ll make personal contacts, find out what the problems are in selecting a research site, and discover how to tie your study to the interests of local scholars. You can study French in France, but you can also study it in Montreal, Martinique, or Madagascar. You can study Spanish in Spain, but you can also study it in Mexico, Bolivia, or Paraguay.

You’d be amazed at the range of language courses available at universities these days: Ulithi, Aymara, Quechua, Nahuatl, Swahili, Turkish, Amharic, Basque, Eskimo, Navajo, Zulu, Hausa, Amoy. . . . If the language you need is not offered in a formal course, try to find an individual speaker of the language (the husband or wife of a foreign student) who would be willing to tutor you in a self-paced course. There are self-paced courses in hundreds of languages available today, many of them on CD, with lots of auditory material.

There are, of course, many languages for which there are no published materials, except perhaps for a dictionary or part of the Judeo-Christian Bible. For those languages, you need to learn how to reduce them to writing quickly so that you can get on with learning them and with fieldwork. To learn how to reduce *any* language to writing, see the tutorial by Oswald Werner (2000a, 2000b, 2001, 2002a, 2002b).

WHEN NOT TO MIMIC

The key to understanding the culture of loggers, lawyers, bureaucrats, schoolteachers, or ethnic groups is to become intimately familiar with their vocabulary. Words are where the cultural action is. My rule about mimicking pronunciation changes, though, if you are studying an ethnic or occupational subculture in your own society and the people in that subculture speak a different dialect of your native language. In this situation, mimicking the local pronunciation will just make you look silly. Even worse, people may think you're ridiculing them.

Building Explicit Awareness

Another important skill in participant observation is what Spradley (1980:55) called **explicit awareness** of the little details in life. Try this experiment: The next time you see someone look at their watch, go right up to them and ask them the time. Chances are they'll look again because when they looked the first time they were not *explicitly aware* of what they saw. Tell them that you are a student conducting a study and ask them to chat with you for a few minutes about how they tell time.

Many people who wear analog watches look at the *relative positions* of the hands, and not at the numbers on the dial. They subtract the current time (the position of the hands now) from the time they have to be somewhere (the image of what the position of the hands will look like at some time in the future), and calculate whether the difference is anything to worry about. They never have to become explicitly aware of the fact that it is 3:10 P.M. People who wear digital watches may be handling the process somewhat differently. We could test that.

Kronenfeld et al. (1972) report an experiment in which informants leaving several different restaurants were asked what the waiters and waitresses (as they were called in those gender-differentiated days) were wearing, and what kind of music was playing. Informants agreed much more about what the waiters were wearing than about what the waitresses were wearing. The hitch: None of the restaurants had waiters, only waitresses.

Informants also provided more detail about the kind of music in restaurants that did not have music than they provided for restaurants that did have music. Kronenfeld et al. speculated that, in the absence of real memories about things they'd seen or heard, informants turned to cultural norms for what must have been there (i.e., "what goes with what") (D'Andrade 1973).

You can test this yourself. Pick out a large lecture hall where a male professor is not wearing a tie. Ask a group of students on their way out of a lecture

hall what color tie their professor was wearing. Or observe a busy store clerk for an hour and count the number of sales she rings up. Then ask her to estimate the number of sales she handled during that hour.

You can build your skills at becoming explicitly aware of ordinary things. Get a group of colleagues together and write separate, detailed descriptions of the most mundane, ordinary things you can think of: making a bed, doing laundry, building a sandwich, shaving (face, legs, underarms), picking out produce at the supermarket, and the like. Then discuss one another's descriptions and see how many details others saw that you didn't and vice versa. If you work carefully at this exercise you'll develop a lot of respect for how complex, and how important, are the details of ordinary life. If you want to see the level of detail you're shooting for here, read Anthony F. C. Wallace's little classic "Driving to Work" (1965). Wallace had made the 17-mile drive from his home to the University of Pennsylvania about 500 times when he drew a map of it, wrote out the details, and extracted a set of rules for his behavior. He was driving a 1962 Volkswagen Beetle in those days. It had 12 major mechanical controls (from the ignition switch to the windshield wiper—yes, there was just one of them, and you had to pull a switch on the instrument panel with your right hand to get it started), all of which had to be handled correctly to get him from home to work safely every day.

Building Memory

Even when we are explicitly aware of things we see, there is no guarantee that we'll remember them long enough to write them down. Building your ability to remember things you see and hear is crucial to successful participant observation research.

Try this exercise: Walk past a store window at a normal pace. When you get beyond it and can't see it any longer, write down all the things that were in the window. Go back and check. Do it again with another window. You'll notice an improvement in your ability to remember little things almost immediately. You'll start to create mnemonic devices for remembering more of what you see. Keep up this exercise until you are satisfied that you can't get any better at it.

Here's another one. Go to a church service, other than one you're used to. Take along two colleagues. When you leave, write up what you each think you saw, in as much detail as you can muster and compare what you've written. Go back to the church and keep doing this exercise until all of you are satisfied that (1) you are all seeing and writing down the same things and (2) you have reached the limits of your ability to recall complex behavioral scenes.

Try this same exercise by going to a church service with which you *are*

familiar and take along several colleagues who are *not*. Again, compare your notes with theirs, and keep going back and taking notes until you and they are seeing and noting the same things. You can do this with any repeated scene that's familiar to you: a bowling alley, a fast-food restaurant, etc. Remember, training your ability to see things *reliably* does not guarantee that you'll see thing *accurately*. But reliability is a necessary but insufficient condition for accuracy. Unless you become at least a reliable instrument of data gathering, you don't stand much of a chance of making valid observations.

Bogdan (1972:41) offers some practical suggestions for remembering details in participant observation. If, for some reason, you can't take notes during an interview or at some event, and you are trying to remember what was said, *don't talk to anyone* before you get your thoughts down on paper. Talking to people reinforces some things you heard and saw at the expense of other things.

Also, when you sit down to write, try to remember things in historical sequence, as they occurred throughout the day. As you write up your notes you will invariably remember some particularly important detail that just pops into memory out of sequence. When that happens, jot it down on a separate piece of paper (or tuck it away in a separate little note file on your word processor) and come back to it later, when your notes reach that point in the sequence of the day.

Another useful device is to draw a map—even a rough sketch will do—of the physical space where you spent time observing and talking to people that day. As you move around the map, you will dredge up details of events and conversations. In essence, let yourself walk through your experience. You can practice all these memory-building skills now and be much better prepared if you decide to do long-term fieldwork later.

Maintaining Naiveté

Try also to develop your skill at being a novice—at being someone who genuinely wants to learn a new culture. This may mean working hard at suspending judgment about some things. David Fetterman made a trip across the Sinai Desert with a group of Bedouins. One of the Bedouins, says Fetterman,

shared his jacket with me to protect me from the heat. I thanked him, of course, because I appreciated the gesture and did not want to insult him. But I smelled like a camel for the rest of the day in the dry desert heat. I thought I didn't need the jacket. . . . I later learned that without his jacket I would have suffered from sunstroke. . . . An inexperienced traveler does not always notice when the temperature climbs above 130 degrees Fahrenheit. By slowing down the evaporation rate, the jacket helped me retain water. (1989:33)

Maintaining your naiveté will come naturally in a culture that's unfamiliar to you, but it's a bit harder to do in your own culture. Most of what you do "naturally" is so automatic that you don't know how to intellectualize it.

If you are like many middle-class Americans, your eating habits can be characterized by the word "grazing"—that is, eating small amounts of food at many, irregular times during the course of a typical day, rather than sitting down for meals at fixed times. Would you have used that kind of word to describe your own eating behavior? Other members of your own culture are often better informants than you are about that culture, and if you really let people teach you, they will.

If you look carefully, though, you'll be surprised at how heterogeneous your culture is and how many parts of it you really know nothing about. Find some part of your own culture that you don't control—an occupational culture, like long-haul trucking, or a hobby culture, like amateur radio—and try to learn it. That's what you did as a child, of course. Only this time, try to intellectualize the experience. Take notes on what you learn about *how to learn*, on what it's like being a novice, and how you think you can best take advantage of the learner's role. Your imagination will suggest a lot of other nooks and crannies of our culture that you can explore as a thoroughly untutored novice.

WHEN NOT TO BE NAIVE

The role of naive novice is not *always* the best one to play. Humility is inappropriate when you are dealing with a culture whose members have a lot to lose by your incompetence. Michael Agar (1973, 1980a) did field research on the life of heroin addicts in New York City. His informants made it plain that Agar's ignorance of their lives wasn't cute or interesting to them.

Even with the best of intentions, Agar could have given his informants away to the police by just by being stupid. Under such circumstances, you shouldn't expect your informants to take you under their wing and teach you how to appreciate their customs. Agar had to learn a lot, and very quickly, to gain credibility with his informants.

There are situations where your expertise is just what's required to build rapport with people. Anthropologists have typed documents for illiterate people in the field and have used other skills (from coaching basketball to dispensing antibiotics) to help people and to gain their confidence and respect. If you are studying highly educated people, you may have to prove that you know a fair amount about research methods before they will deal with you. Agar (1980b:58) once studied an alternative lifestyle commune and was asked by a biochemist who was living there: "Who are you going to use as a control

group?” In my study of ocean scientists (Bernard 1974), several informants asked me what computer programs I was going to use to do a factor analysis of my data.

Building Writing Skills

The ability to write comfortably, clearly, and often is one of the most important skills you can develop as a participant observer. Ethnographers who are not comfortable as writers produce few field notes and little published work. If you have any doubts about your ability to pound out thousands of words, day in and day out, then try to build that skill now, before you go into the field for an extended period.

The way to build that skill is to team up with one or more colleagues who are also trying to build their expository writing ability. Set concrete and regular writing tasks for yourselves and criticize one another's work on matters of clarity and style. There is nothing trivial about this kind of exercise. If you think you need it, do it.

Good writing skills will carry you through participant observation fieldwork, writing a dissertation and, finally, writing for publication. Don't be afraid to write clearly and compellingly. The worst that can happen is that someone will criticize you for “popularizing” your material. I think ethnographers should be criticized if they take the exciting material of real people's lives and turn it into deadly dull reading.

Hanging Out, Gaining Rapport

It may sound silly, but just **hanging out** is a skill, and until you learn it you can't do your best work as a participant observer. Remember what I said at the beginning of this chapter: Participant observation is *a strategic* method that lets you learn what you want to learn and apply all the data collection methods that you may want to apply.

When you enter a new field situation, the temptation is to ask a lot of questions in order to learn as much as possible as quickly as possible. There are many things that people can't or won't tell you in answer to questions. If you ask people too quickly about the sources of their wealth, you are likely to get incomplete data. If you ask too quickly about sexual liaisons, you may get thoroughly unreliable responses.

Hanging out builds trust, or **rapport**, and trust results in ordinary conversation and ordinary behavior in your presence. Once you know, from hanging out, exactly what you want to know more about, and once people trust you

not to betray their confidence, you'll be surprised at the direct questions you can ask.

In his study of Cornerville (Boston's heavily Italian American neighborhood called North End), William Foote Whyte wondered whether "just hanging on the street corner was an active enough process to be dignified by the term 'research.' Perhaps I should ask these men questions," he thought. He soon realized that "one has to learn when to question and when not to question as well as what questions to ask" (1989:78).

Philip Kilbride studied child abuse in Kenya. He did a survey and focused ethnographic interviews, but "by far the most significant event in my research happened as a byproduct of participatory 'hanging out', being always in search of case material." While visiting informants one day, Kilbride and his wife saw a crowd gathering at a local secondary school. It turned out that a young mother had thrown her baby into a pit latrine at the school. The Kilbrides offered financial assistance to the young mother and her family in exchange for "involving ourselves in their . . . misfortune." The event that the Kilbrides had witnessed became the focus for a lot of their research activities in the succeeding months (Kilbride 1992:190).

THE ETHICAL DILEMMA OF RAPPORT

Face it: "Gaining rapport" is a euphemism for impression management, one of the "darker arts" of fieldwork, in Harry Wolcott's apt phrase (2005:chap. 6). E. E. Evans-Pritchard, the great British anthropologist, made clear in 1937 how manipulative the craft of ethnography really is. He was doing fieldwork with the Azande of Sudan and wanted to study their rich tradition of witchcraft. Even with his long-term fieldwork and command of the Azande language, Evans-Pritchard couldn't get people to open up about witchcraft, so he decided to "win the good will of one or two practitioners and to persuade them to divulge their secrets in strict confidence" (1958 [1937]:151). Strict confidence? He was planning on writing a book about all this.

Progress was slow, and while he felt that he could have "eventually wormed out all their secrets" he hit on another idea: His personal servant, Kamanga, was initiated into the local group of practitioners and "became a practising witch-doctor" under the tutelage of a man named Badobo (*ibid.*). With Badobo's full knowledge, Kamanga reported every step of his training to his employer. In turn, Evans-Pritchard used the information "to draw out of their shells rival practitioners by playing on their jealousy and vanity."

Now, Badobo knew that anything he told Kamanga would be tested with rival witch doctors. Badobo couldn't lie to Kamanga, but he could certainly

withhold the most secret material. Evans-Pritchard analyzed the situation carefully and pressed on. Once an ethnographer is “armed with preliminary knowledge,” he said, “nothing can prevent him from driving deeper and deeper the wedge if he is interested and persistent” (ibid.:152).

Still, Kamanga’s training was so slow that Evans-Pritchard nearly abandoned his inquiry into witchcraft. Providence intervened. A celebrated witch doctor, named Bögwözu, showed up from another district and Evans-Pritchard offered him a very high wage if he’d take over Kamanga’s training. Evans-Pritchard explained to Bögwözu that he was “tired of Badobo’s wiliness and extortion,” and that he expected his generosity to result in Kamanga learning all the tricks of the witch doctor’s trade (ibid.).

But the really cunning part of Evans-Pritchard’s scheme was that he continued to pay Badobo to tutor Kamanga. He *knew* that Badobo would be jealous of Bögwözu and would strive harder to teach Kamanga more about witch-doctoring. Here is Evans-Pritchard going on about his deceit and the benefits of this tactic for ethnographers:

The rivalry between these two practitioners grew into bitter and ill-concealed hostility. Bögwözu gave me information about medicines and magical rites to prove that his rival was ignorant of the one or incapable in the performance of the other. Badobo became alert and showed himself no less eager to demonstrate his knowledge of magic to both Kamanga and to myself. They vied with each other to gain ascendancy among the local practitioners. Kamanga and I reaped a full harvest in this quarrel, not only from the protagonists themselves but also from other witch-doctors in the neighborhood, and even from interested laymen. (ibid.:153)

Objectivity

Finally, **objectivity** is a skill, like language fluency, and you can build it if you work at it. Some people build more of it, others less. More is better.

If an objective measurement is one made by a robot—that is, a machine that is not prone to the kind of measurement error that comes from having opinions and memories—then no human being can ever be completely objective. We can’t rid ourselves of our experiences, and I don’t know anyone who thinks it would be a good idea even to try.

We can, however, become aware of our experiences, our opinions, our values. We can hold our field observations up to a cold light and ask whether we’ve seen what we wanted to see, or what is really out there. The goal is not for us, as humans, to become objective machines; it is for us to achieve objective—that is, accurate—knowledge by transcending our biases. No fair pointing out that this is impossible. Of course, it’s impossible to do com-

pletely. But it's not impossible to do at all. Priests, social workers, clinical psychologists, and counselors suspend their own biases all the time, more or less, in order to listen hard and give sensible advice to their clients.

Laurie Krieger, an American woman doing fieldwork in Cairo, studied physical punishment against women. She learned that wife beatings were less violent than she had imagined and that the act still sickened her. Her reaction brought out a lot of information from women who were recent recipients of their husbands' wrath. "I found out," she says, "that the biased outlook of an American woman and a trained anthropologist was not always disadvantageous, as long as I was aware of and able to control the expression of my biases" (Kreiger 1986:120).

Colin Turnbull held objective knowledge as something to be pulled from the thicket of subjective experience. Fieldwork, said Turnbull, involves a self-conscious review of one's own ideas and values—one's self, for want of any more descriptive term. During fieldwork you "reach inside," he observed, and give up the "old, narrow, limited self, discovering the new self that is right and proper in the new context." We use the field experience, he said, "to know ourselves more deeply by conscious subjectivity." In this way, he concluded, "the ultimate goal of objectivity is much more likely to be reached and our understanding of other cultures that much more profound" (Turnbull 1986:27). When he was studying the Ik of Uganda, he saw parents goad small children into touching fire and then laughing at the result. It took everything he had, he once told me, to transcend his biases, but he managed (see Turnbull 1972).

Many phenomenologists see objective knowledge as the goal of participant observation. Danny Jorgensen, for example, advocates complete immersion and **becoming the phenomenon** you study. "Becoming the phenomenon," Jorgensen says, "is a participant observational strategy for penetrating to and gaining experience of a form of human life. It is an objective approach insofar as it results in the accurate, detailed description of the insiders' experience of life" (Jorgensen 1989:63). In fact, many ethnographers have become cab drivers or exotic dancers, jazz musicians, or members of satanic cults, in order to do participant observation fieldwork.

If you use this strategy of full immersion, Jorgensen says, you must be able to switch back and forth between the insiders' view and that of an analyst. To do that—to maintain your objective, analytic abilities—Jorgensen suggests finding a colleague with whom you can talk things over regularly. That is, give yourself an outlet for discussing the theoretical, methodological, and emotional issues that inevitably come up in full participation field research. It's good advice.

OBJECTIVITY AND NEUTRALITY

Objectivity does not mean (and has never meant) **value neutrality**. No one asks Cultural Survival, Inc. to be neutral in documenting the violent obscenities against indigenous peoples of the world. No one asks Amnesty International to be neutral in its effort to document state-sanctioned torture. We recognize that the power of the documentation is in its objectivity, in its chilling irrefutability, not in its neutrality.

Claire Sterk, an ethnographer from the Netherlands, has studied prostitutes and intravenous drug users in mostly African American communities in New York City and Newark, New Jersey. Sterk was a trusted friend and counselor to many of the women with whom she worked. In one 2-month period in the late 1980s, she attended the funerals of seven women she knew who had died of AIDS. She felt that “every researcher is affected by the work he or she does. One cannot remain neutral and uninvolved; even as an outsider, the researcher is part of the community” (Sterk 1989:99, 1999).

At the end of his second year of research on street life in El Barrio, Phillipe Bourgois’s friends and informants began telling him about their experiences as gang rapists. Bourgois’s informants were in their mid- to late 20s then, and the stories they told were of things they’d done as very young adolescents, more than a decade earlier. Still, Bourgois says, he felt betrayed by people whom he had come to like and respect. Their “childhood stories of violently forced sex,” he says, “spun me into a personal depression and a research crisis” (1995:205).

In *any* long-term field study, be prepared for some serious tests of your ability to remain a dispassionate observer. Hortense Powdermaker (1966) was once confronted with the problem of knowing that a lynch mob was preparing to go after a particular black man. She was powerless to stop the mob and fearful for her own safety.

I have never grown accustomed to seeing people ridicule the handicapped, though I see it every time I’m in rural Mexico and Greece, and I recall with horror the death of a young man on one of the sponge diving boats I sailed with in Greece. I knew the rules of safe diving that could have prevented that death; so did all the divers and the captains of the vessels. They ignored those rules at terrible cost. I wanted desperately to *do* something, but there was nothing anyone could do. My lecturing them at sea about their unsafe diving practices would not have changed their behavior. That behavior was driven, as I explained in chapter 2, by structural forces and the technology—the boats, the diving equipment—of their occupation. By suspending active judgment of their behavior, I was able to record it. “Suspending active judgment” does not mean that I eliminated my bias or that my feelings about their behavior

changed. It meant only that I kept the bias to myself while I was recording their dives.

OBJECTIVITY AND INDIGENOUS RESEARCH

Objectivity gets its biggest test in **indigenous research**—that is, when you study your own culture. Barbara Meyerhoff worked in Mexico when she was a graduate student. Later, in the early 1970s, when she became interested in ethnicity and aging, she decided to study elderly Chicanos. The people she approached kept putting her off, asking her “Why work with us? Why don’t you study your own kind?” Meyerhoff was Jewish. She had never thought about studying her own kind, but she launched a study of poor, elderly Jews who were on public assistance. She agonized about what she was doing and, as she tells it, never resolved whether it was anthropology or a personal quest.

Many of the people she studied were survivors of the Holocaust. “How, then, could anyone look at them dispassionately? How could I feel anything but awe and appreciation for their mere presence? . . . Since neutrality was impossible and idealization undesirable, I decided on striving for balance” (Meyerhoff 1989:90).

There is no final answer on whether it’s good or bad to study your own culture. Plenty of people have done it, and plenty of people have written about what it’s like to do it. On the plus side, you’ll know the language and you’ll be less likely to suffer from culture shock. On the minus side, it’s harder to recognize cultural patterns that you live every day and you’re likely to take a lot of things for granted that an outsider would pick up right away.

If you are going to study your own culture, start by reading the experiences of others who have done it so you’ll know what you’re facing in the field (Messerschmidt 1981; Stephenson and Greer 1981; Fahim 1982; Altorki and El-Solh 1988). (See the section on native ethnographies in chapter 17 for more about indigenous research.)

Gender, Parenting, and Other Personal Characteristics

By the 1930s, Margaret Mead had already made clear the importance of gender as a variable in data collection (see Mead 1986). Gender has at least two consequences: (1) It limits your access to certain information; (2) It influences how you perceive others.

In all cultures, you can’t ask people certain questions because you’re a [woman] [man]. You can’t go into certain areas and situations because you’re a [woman] [man]. You can’t watch this or report on that because you’re a

[woman] [man]. Even the culture of social scientists is affected: Your credibility is diminished or enhanced with your colleagues when you talk about a certain subject because you're a [woman] [man] (Scheper-Hughes 1983; Golde 1986; Whitehead and Conaway 1986; Altorki and El-Solh 1988; Warren 1988).

Sara Quandt, Beverly Morris, and Kathleen DeWalt spent months investigating the nutritional strategies of the elderly in two rural Kentucky counties (Quandt et al. 1997). According to DeWalt, the three women researchers spent months, interviewing key informants, and never turned up a word about the use of alcohol. "One day," says DeWalt,

the research team traveled to Central County with Jorge Uquillas, an Ecuadorian sociologist who had expressed an interest in visiting the Kentucky field sites. One of the informants they visited was Mr. B, a natural storyteller who had spoken at length about life of the poor during the past sixty years. Although he had been a great source of information about use of wild foods and recipes for cooking game he had never spoken of drinking or moonshine production.

Within a few minutes of entering his home on this day, he looked at Jorge Uquillas, and said "Are you a drinking man?" (Beverly whipped out the tape recorder and switched it on.) Over the next hour or so, Mr. B talked about community values concerning alcohol use, the problems of drunks and how they were dealt with in the community, and provided a number of stories about moonshine in Central County. The presence of another man gave Mr. B the opportunity to talk about issues he found interesting, but felt would have been inappropriate to discuss with women. (DeWalt et al. 1998:280)

On the other hand, feminist scholars have made it clear that gender is a negotiated idea. What you can and can't do if you are a man or a woman is more fixed in some cultures than in others, and in all cultures there is lots of individual variation in gender roles. While men or women may be expected to be this way or that way in any given place, the variation in male and female attitudes and behaviors within a culture can be tremendous.

All participant observers confront their personal limitations and the limitations imposed on them by the culture they study. When she worked at the Thule relocation camp for Japanese Americans during World War II, Rosalie Wax did not join any of the women's groups or organizations. Looking back after more than 40 years, Wax concluded that this was just poor judgment.

I was a university student and a researcher. I was not yet ready to accept myself as a total person, and this limited my perspective and my understanding. Those of us who instruct future field workers should encourage them to understand and value their full range of being, because only then can they cope intelligently with the range of experience they will encounter in the field. (Wax 1986:148)

Besides gender, we have learned that being a parent helps you talk to people about certain areas of life and get more information than if you were not a parent. My wife and I arrived on the island of Kalymnos, Greece, in 1964 with a 2-month-old baby. As Joan Cassell says, children are a “guarantee of good intentions” (1987:260), and wherever we went, the baby was the conversation opener. But be warned: Taking children into the field can place them at risk. (More on health risks below. And for more about the effects of fieldwork on children who accompany researchers, see Butler and Turner [1987].)

Being divorced has its costs. Nancie González found that being a divorced mother of two young sons in the Dominican Republic was just too much. “Had I to do it again,” she says, “I would invent widowhood with appropriate rings and photographs” (1986:92).

Even height may make a difference: Alan Jacobs once told me he thought he did better fieldwork with the Maasai because he’s 6’5” than he would have if he’d been, say, an average-sized 5’10”.

Personal characteristics make a difference in fieldwork. Being old or young lets you into certain things and shuts you out of others. Being wealthy lets you talk to certain people about certain subjects and makes others avoid you. Being gregarious makes some people open up to you and makes others shy away. There is no way to eliminate the “personal equation” in participant observation fieldwork, or in any other scientific data-gathering exercise for that matter, without sending robots out to do the work. Of course, the robots would have their own problems. In all sciences, the personal equation (the influence of the observer on the data) is a matter of serious concern and study (Romney 1989).

Sex and Fieldwork

It is unreasonable to assume that single, adult fieldworkers are all celibate, yet the literature on field methods was nearly silent on this topic for many years. When Evans-Pritchard was a student, just about to head off for Central Africa, he asked his major professor for advice. “Seligman told me to take ten grains of quinine every night and keep off women” (Evans-Pritchard 1973:1). As far as I know, that’s the last we heard from Evans-Pritchard on the subject.

Colin Turnbull (1986) tells us about his affair with a young Mbuti woman, and Dona Davis (1986) discusses her relationship with an engineer who visited the Newfoundland village where she was doing research on menopause. In Turnbull’s case, he had graduated from being an asexual child in Mbuti culture to being a youth and was expected to have sexual relations. In Davis’s case, she was expected not to have sexual relations, but she also learned that

she was not bound by the expectation. In fact, Davis says that “being paired off” made women more comfortable with her because she was “simply breaking a rule everyone else broke” (1986:254).

With changing sexual mores in our late industrial society, anthropologists have become more open about the topic of sex and fieldwork. Several anthologies have been published in which researchers discuss their own sexual experiences during participant observation fieldwork (Kulick and Willson 1995; Lewin and Leap 1996; Markowitz and Ashkenazi 1999). Proscriptions against sex in fieldwork are silly, because they don’t work. But understand that this is one area that people everywhere take very seriously.

The rule on sexual behavior in the field is this: Do nothing that you can’t live with, both professionally and personally. This means that you have to be even more conscious of any fallout, for you and for your partner, than you would in your own community. Eventually, you will be going home. How will that affect your partner’s status?

Surviving Fieldwork

The title of this section is the title of an important book by Nancy Howell (1990). All researchers—whether they are anthropologists, epidemiologists, or social psychologists—who expect to do fieldwork in developing nations should read that book. Howell surveyed 204 anthropologists about illnesses and accidents in the field, and the results are sobering. The maxim that “anthropologists are otherwise sensible people who don’t believe in the germ theory of disease” is apparently correct (Rappaport 1990).

One hundred percent of anthropologists who do fieldwork in south Asia reported being exposed to malaria, and 41% reported contracting the disease. Eighty-seven percent of anthropologists who work in Africa reported exposure, and 31% reported having had malaria. Seventy percent of anthropologists who work in south Asia reported having had some liver disease.

Among all anthropologists, 13% reported having had hepatitis A. I was hospitalized for 6 weeks for hepatitis A in 1968 and spent most of another year recovering. Glynn Isaac died of hepatitis B at age 47 in 1985 after a long career of archeological fieldwork in Africa. Typhoid fever is also common among anthropologists, as are amoebic dysentery, giardia, ascariasis, hookworm, and other infectious diseases.

Accidents have injured or killed many fieldworkers. Fei Xiaotong, a student of Malinowski’s, was caught in a tiger trap in China in 1935. The injury left him an invalid for 6 months. His wife died in her attempt to go for help. Michelle Zimbalist Rosaldo was killed in a fall in the Philippines in 1981.

Thomas Zwickler, a graduate student at the University of Pennsylvania, was killed by a bus on a rural road in India in 1985. He was riding a bicycle when he was struck. Kim Hill was accidentally hit by an arrow while out with an Ache hunting party in Paraguay in 1982 (Howell 1990: *passim*).

Five members of a Russian-American team of researchers on social change in the Arctic died in 1995 when their *umiak* (a traditional, walrus-hided Eskimo boat) was overturned by a whale (see Broadbent 1995). The researchers included three Americans (two anthropologists—Steven McNabb and Richard Condon—and a psychiatrist—William Richards), and two Russians (one anthropologist—Alexander Pika—and the chief Eskimo ethnographic consultant to the project—Boris Mumikhpykak). Nine Eskimo villagers also perished in that accident. I've had my own unpleasant brushes with fate and I know many others who have had very, very close calls.

What can you do about the risks? Get every inoculation you need before you leave, not just the ones that are required by the country you are entering. Check your county health office for the latest information from the Centers for Disease Control about illnesses prevalent in the area you're going to. If you go into an area that is known to be malarial, take a full supply of antimalarial drugs with you so you don't run out while you're out in the field.

When people pass around a gourd full of *chicha* (beer made from corn) or *pulque* (beer made from cactus sap) or palm wine, decline politely and explain yourself if you have to. You'll probably insult a few people, and your protests won't always get you off the hook, but even if you only lower the number of times you are exposed to disease, you lower your risk of contracting disease.

After being very sick in the field, I learned to carry a supply of bottled beer with me when I'm going to visit a house where I'm sure to be given a gourd full of local brew. The gift of bottled beer is generally appreciated and heads off the embarrassment of having to turn down a drink I'd rather not have. It also makes plain that I'm not a teetotaler. Of course, if you *are* a teetotaler, you've got a ready-made get-out.

If you do fieldwork in a remote area, consult with physicians at your university hospital for information on the latest blood-substitute technology. If you are in an accident in a remote area and need blood, a nonperishable blood substitute can buy you time until you can get to a clean blood supply. Some fieldworkers carry a supply of sealed hypodermic needles with them in case they need an injection. Don't go anywhere without medical insurance and don't go to developing countries without evacuation insurance. It costs about \$60,000 to evacuate a person by jet from central Africa to Paris or Frankfurt. It costs about \$50 a month for insurance to cover it.

Fieldwork in remote areas isn't for everyone, but if you're going to do it, you might as well do it as safely as possible. Candice Bradley is a Type-I

diabetic who does long-term fieldwork in western Kenya. She takes her insulin, glucagon, blood-testing equipment, and needles with her. She arranges her schedule around the predictable, daily fluctuations in her blood-sugar level. She trains people on how to cook for her and she lays in large stocks of diet drinks so that she can function in the relentless heat without raising her blood sugars (Bradley 1997:4–7).

With all this, Bradley still had close calls—near blackouts from hypoglycemia—but her close calls are no more frequent than those experienced by other field researchers who work in similarly remote areas. The rewards of foreign fieldwork can be very great, but so are the risks.

The Stages of Participant Observation

In what follows, I will draw on three sources of data: (1) a review of the literature on field research; (2) conversations with colleagues during the last 40 years, specifically about their experiences in the field; and (3) 5 years of work, with the late Michael Kenny, directing National Science Foundation field schools in cultural anthropology and linguistics.

During our work with the field schools (1967–1971), Kenny and I developed an outline of **researcher response** in participant observation fieldwork. Those field schools were 10 weeks long and were held each summer in central Mexico, except for one that was held in the interior of the Pacific Northwest. In Mexico, students were assigned to Nāhñu-speaking communities in the vicinity of Ixmiquilpan, Mexico. In the Northwest field school, students were assigned to small logging and mining communities in the Idaho panhandle. In Mexico, a few students did urban ethnography in the regional capital of Pachuca, while in the Northwest field school, a few students did urban ethnography in Spokane, Washington.

What Kenny and I found so striking was that the stages we identified in the 10-week field experiences of our students were the same across all these places. Even more interesting—to us, anyway—was that the experiences our students had during those 10-week stints as participant observers apparently had exact analogs in our own experiences with yearlong fieldwork.

1. Initial Contact

During the initial contact period, many long-term fieldworkers report experiencing a kind of euphoria as they begin to move about in a new culture. It shouldn't come as any surprise that people who are attracted to the idea of living in a new culture are delighted when they begin to do so.

But not always. Here is Napoleon Chagnon's recollection of his first encounter with the Yanomami: "I looked up and gasped when I saw a dozen burly, naked, sweaty, hideous men staring at us down the shafts of their drawn arrows! . . . had there been a diplomatic way out, I would have ended my fieldwork then and there" (Chagnon 1983:10–11).

The desire to bolt and run is more common than we have admitted in the past. Charles Wagley, who would become one of our discipline's most accomplished ethnographers, made his first field trip in 1937. A local political chief in Totonicapán, Guatemala, invited Wagley to tea in a parlor overlooking the town square. The chief's wife and two daughters joined them. While they were having their tea, two of the chief's aides came in and hustled everyone off to another room. The chief explained the hurried move to Wagley:

He had forgotten that an execution by firing squad of two Indians, "nothing but vagrants who had robbed in the market," was to take place at five P.M. just below the parlor. He knew that I would understand the feelings of ladies and the grave problem of trying to keep order among brutes. I returned to my ugly *pensión* in shock and spent a night without sleep. I would have liked to have returned as fast as possible to New York. (Wagley 1983:6)

Finally, listen to Rosalie Wax describe her encounter with the Arizona Japanese internment camp that she studied during World War II. When she arrived in Phoenix it was 110°. Later that day, after a bus ride and a 20-mile ride in a GI truck, across a dusty landscape that "looked like the skin of some cosmic reptile," with a Japanese American who wouldn't talk to her, Wax arrived at the Gila camp. By then it was 120°. She was driven to staff quarters, which was an army barracks divided into tiny cells, and abandoned to find her cell by a process of elimination.

It contained four dingy and dilapidated articles of furniture: an iron double bedstead, a dirty mattress (which took up half the room), a chest of drawers, and a tiny writing table—and it was hotter than the hinges of Hades. . . . I sat down on the hot mattress, took a deep breath, and cried. . . . Like some lost two-year-old, I only knew that I was miserable. After a while, I found the room at the end of the barrack that contained two toilets and a couple of wash basins. I washed my face and told myself I would feel better the next day. I was wrong. (Wax 1971:67)

2. *Culture Shock*

Even among those fieldworkers who have a pleasant experience during their initial contact period (and many do), almost all report experiencing some form of depression and shock soon thereafter—usually within a few weeks. (The

term “culture shock,” by the way, was introduced in 1960 by an anthropologist, Kalervo Oberg.) One kind of shock comes as the novelty of the field site wears off and there is this nasty feeling that research has to get done. Some researchers (especially those on their first field trip) may also experience feelings of anxiety about their ability to collect good data.

A good response at this stage is to do highly task-oriented work: making maps, taking censuses, doing household inventories, collecting genealogies, and so on. Another useful response is to make clinical, methodological field notes about your feelings and responses in doing participant observation fieldwork.

Another kind of shock is to the culture itself. **Culture shock** is an uncomfortable stress response and must be taken very seriously. In serious cases of culture shock, nothing seems right. You may find yourself very upset at a lack of clean toilet facilities, or people’s eating habits, or their child-rearing practices. The prospect of having to put up with the local food for a year or more may become frightening. You find yourself focusing on little annoyances—something as simple as the fact that light switches go side to side rather than up and down may upset you.

This last example is not fanciful, by the way. It happened to a colleague of mine. When I first went to work with the Nāhñu in 1962, men would greet me by sticking out their right hand. When I tried to grab their hand and shake it, they deftly slid their hand to my right so that the back of their right hand touched the back of my right hand. I became infuriated that men didn’t shake hands the way “they’re supposed to.” You may find yourself blaming everyone in the culture, or the culture itself, for the fact that your informants don’t keep appointments for interviews.

Culture shock commonly involves a feeling that people really don’t want you around (this may, in fact, be the case). You feel lonely and wish you could find someone with whom to speak your native language. Even with a spouse in the field, the strain of using another language day after day, and concentrating hard so that you can collect data in that language, can be emotionally wearing.

A common personal problem in field research is not being able to get any privacy. Many people across the world find the Anglo-Saxon notion of privacy grotesque. When we first went out to the island of Kalymnos in Greece in 1964, Carole and I rented quarters with a family. The idea was that we’d be better able to learn about family dynamics that way. Women of the household were annoyed and hurt when my wife asked for a little time to be alone. When I came home at the end of each day’s work, I could never just go to my family’s room, shut the door, and talk to Carole about my day, or hers, or our new

baby's. If I didn't share everything with the family we lived with during waking hours, they felt rejected.

After about 2 months of this, we had to move out and find a house of our own. My access to data about intimate family dynamics was curtailed. But it was worth it because I felt that I'd have had to abort the whole trip if I had to continue living in what my wife and I felt was a glass bowl all the time. As it turns out, there is no word for the concept of privacy in Greek. The closest gloss translates as "being alone," and connotes loneliness.

I suspect that this problem is common to all English-speaking researchers who work in developing countries. Here's what M. N. Srinivas, himself from India, wrote about his work in the rural village of Ramapura, near Mysore:

I was never left alone. I had to fight hard even to get two or three hours absolutely to myself in a week or two. My favorite recreation was walking to the nearby village of Kere where I had some old friends, or to Hogur which had a weekly market. But my friends in Ramapura wanted to accompany me on my walks. They were puzzled by my liking for solitary walks. Why should one walk when one could catch a bus, or ride on bicycles with friends. I had to plan and plot to give them the slip to go out by myself. On my return, however, I was certain to be asked why I had not taken them with me. They would have put off their work and joined me. (They meant it.) I suffered from social claustrophobia as long as I was in the village and sometimes the feeling became so intense that I just had to get out. (1979:23)

Culture shock subsides as researchers settle in to the business of gathering data on a daily basis, but it doesn't go away because the sources of annoyance don't go away.

Unless you are one of the very rare people who truly go native in another culture, you will cope with culture shock, not eliminate it. You will remain conscious of things annoying you, but you won't feel like they are crippling your ability to work. Like Srinivas, when things get too intense, you'll have the good sense to leave the field site for a bit rather than try to stick it out. (For more about culture shock, see Furnham and Bochner 1986, Mumford 1998, and Bochner 2000.)

3. *Discovering the Obvious*

In the next phase of participant observation, researchers settle into collecting data on a more or less systematic basis (see Kirk and Miller 1986). This is sometimes accompanied by an interesting personal response—a sense of discovery, where you feel as if informants are finally letting you in on the "good stuff" about their culture. Much of this "good stuff" will later turn out

to be commonplace. You may “discover,” for example, that women have more power in the community than meets the eye or that there are two systems for dispute settlement—one embodied in formal law and one that works through informal mechanisms.

Sometimes, a concomitant to this feeling of discovery is a feeling of being in control of dangerous information and a sense of urgency about protecting informants’ identities. You may find yourself going back over your field notes, looking for places that you might have lapsed and identified an informant, and making appropriate changes. You may worry about those copies of field notes you have already sent home and even become a little worried about how well you can trust your major professor to maintain the privacy of those notes.

This is the stage of fieldwork when you hear anthropologists start talking about “their” village, and how people are, at last, “letting them in” to the secrets of the culture. The feeling has its counterpart among all long-term participant observers. It often spurs researchers to collect more and more data; to accept every invitation, by every informant, to every event; to fill the days with observation, and to fill the nights with writing up field notes. Days off become unthinkable, and the sense of discovery becomes more and more intense.

This is the time to take a serious break.

4. The Break

The mid-fieldwork break, which usually comes after 3 or 4 months, is a crucial part of the overall participant observation experience for long-term researchers. It’s an opportunity to get some distance, both physical and emotional, from the field site. It gives you a chance to put things into perspective, think about what you’ve got so far, and what you need to get in the time remaining. Use this time to collect data from regional or national statistical services; to visit with colleagues at the local university and discuss your findings; to visit other communities in other parts of the country. And be sure to leave some time to just take a vacation, without thinking about research at all.

Your informants also need a break from you. “Anthropologists are uncomfortable intruders no matter how close their rapport,” wrote Charles Wagley. “A short respite is mutually beneficial. One returns with objectivity and human warmth restored. The anthropologist returns as an old friend” who has gone away and returned, and has thereby demonstrated his or her genuine interest in a community (Wagley 1983:13). Everyone needs a break.

5. Focusing

After the break, you will have a better idea of exactly what kinds of data you are lacking, and your sense of the problem will also come more sharply

into focus. The reason to have a formally prepared design statement *before* you go to the field is to tell you what you should be looking for. Nevertheless, even the most focused research design will have to be modified in the field. In some cases, you may find yourself making radical changes in your design, based on what you find when you get to the field and spend several months actually collecting data.

There is nothing wrong or unusual about this, but new researchers sometimes experience anxiety over making any major changes. The important thing at this stage is to focus the research and use your time effectively rather than agonizing over how to save components of your original design.

6. Exhaustion, the Second Break, and Frantic Activity

After 7 or 8 months, some participant observers start to think that they have exhausted their informants, both literally and figuratively. That is, they may become embarrassed about continuing to ask informants for more information. Or they may make the supreme mistake of believing that their informants have no more to tell them. The reason this is such a mistake, of course, is that the store of cultural knowledge in any culturally competent person is enormous—far more than anyone could hope to extract in a year or two.

At this point, another break is usually a good idea. You'll get another opportunity to take stock, order your priorities for the time remaining, and see both how much you've done and how little. The realization that, in fact, informants have a great deal more to teach them, and that they have precious little time left in the field, sends many investigators into a frenetic burst of activity during this stage.

7. Leaving the Field

The last stage of participant observation is leaving the field. When should you leave? Steven Taylor, a sociologist at the Center for Human Policy, says that when he starts to get bored writing field notes, he knows it's time to close down and go home. Taylor recognizes that writing field notes is time consuming and tedious, but it's exciting, too, when you're chasing down information that plugs directly into your research effort (Taylor 1991:243). When it stops being exciting, it's time to leave the field.

Don't neglect this part of the process. Let people know that you are leaving and tell them how much you appreciate their help. The ritual of leaving a place in a culturally appropriate way will make it possible for you to go back and even to send others.

Participant observation is an intensely intimate and personal experience.

People who began as your informants may become your friends as well. In the best of cases, you come to trust that they will not deceive you about their culture, and they come to trust you not to betray them—that is, not to use your intimate knowledge of their lives to hurt them. (You can imagine the worst of cases.) There is often a legitimate expectation on both sides that the relationship may be permanent, not just a 1-year fling.

For many long-term participant observation researchers, there is no final leaving of “the field.” I’ve been working with some people, on and off, for 40 years. Like many anthropologists who work in Latin America, I’m godparent to a child of my closest research collaborator. From time to time, people from Mexico or from Greece will call my house on the phone, just to say “hi” and to keep the relationship going.

Or their children, who happen to be doing graduate work at a university in the United States, will call and send their parents’ regards. They’ll remind you of some little event they remember when they were 7 or 8 and you came to their parents’ house to do some interviewing and you spilled your coffee all over yourself as you fumbled with your tape recorder. People remember the darndest things. You’d better be ready when it happens.

Many fieldworkers have been called on to help the children of their informants get into a college or university. This is the sort of thing that happens 20 years after you’ve “left” the field. The fact is, participant observation fieldwork can be a lifetime commitment. As in all aspects of ordinary life, you have to learn to choose your relationships well. Don’t be surprised if you make a few mistakes.

The Front-Edge: Combining Methods

More and more researchers these days, across the social sciences, have learned what a powerful method powerful participant observation is at all stages of the research process. The method stands on its own, but it is also increasingly part of a mixed-method strategy, as researchers combine qualitative and quantitative data to answer questions of interest.

Laura Miller (1997) used a mix of ethnographic and survey methods to study gender harassment in the U.S. Army. Keeping women out of jobs that have been traditionally reserved for men is *gender* harassment; asking women for sex in return for a shot at one of those jobs is *sexual* harassment. (Gender harassment need not involve sexual harassment, or vice versa.)

Miller spent nearly 2 years collecting data at eight army posts and at two training centers in the United States where war games are played out on simulated battlefields. She lived in Somalia with U.S. Army personnel for 10 days,

in Macedonia for a week, and in Haiti for 6 days during active military operations in those countries. Within the context of participant observation, she did unstructured interviewing, in-depth interviewing, and group interviewing. Her group interviews were spontaneous: over dinner with a group of high-ranking officers; sitting on her bunk at night, talking to her roommates; in vehicles, bouncing between research sites, with the driver, guide, protocol officer, translator, and guard (Miller, personal communication).

It turns out that “forms of gender harassment” in the U.S. Army is one of those cultural domains that people recognize and think about, but for which people have no ready list in their heads. You can’t just ask people: “List the kinds of gender harassment.” From her ethnographic interviews, though, Miller was able to derive what she felt was just such a list, including:

1. *resistance* to authority (hostile enlisted men ignore orders from women officers);
2. *constant scrutiny* (men pick up on every mistake that women make and use those mistakes to criticize the abilities of women in general);
3. *gossip and rumors* (women who date many men are labeled “sluts,” women who don’t date at all are labeled “dykes,” and any woman can easily be unjustly accused of “sleeping her way to the top”);
4. *outright sabotage* of women’s tools and equipment on work details; and
5. *indirect threats* against women’s safety (talking about how women would be vulnerable to rape if they were to go into combat).

This list emerges from qualitative research—hanging out, talking to people and gaining their trust, and generally letting people know that you’re in for the long haul with them. If you are trying to develop programs to correct things that are wrong with a program, then this list, derived entirely from participant observation, is enough. An education program to counter gender harassment against women in the U.S. Army must include something about each of the problems that Miller identified.

Although ethnographic methods are enough to *identify* the problems and processes—the what and the how of culture—ethnography can’t tell you *how much* each problem and process counts. Yes, enlisted army men can and do sabotage army women’s tools and equipment on occasion. How often? Ethnography can’t help with that one. Yes, men do sometimes resist the authority of women officers. How often? Ethnography can’t help there, either.

Fortunately, Miller also collected questionnaire data—from a quota sample of 4,100 men and women, Whites and Blacks, officers and enlisted personnel. In those data, 19% of enlisted men and 18% of male noncommissioned officers (like sergeants) said that women should be treated exactly like men and should serve in the combat units just like men, while just 6% of enlisted women and 4% of female noncommissioned officers agreed with this senti-

ment. You might conclude, Miller says, that men are more supportive than women are of equality for women in combat roles. Some men with whom Miller spoke, however, said that women should be given the right to serve in combat *so that, once and for all, everyone will see that women can't cut it.*

Are men really what Miller called “hostile proponents” of equality for women? Could that be why the statistics show so many more men in favor of women serving in combat units? Miller went back to her questionnaire data: About 20% of men in her survey said that women should be assigned to combat units just like men were—but almost to a man they also said that putting women into combat units would reduce the military’s effectiveness.

In other words, the numerical analysis showed that Miller’s concept of “hostile proponent of equality” was correct. This subtle concept advances our understanding considerably of how gender harassment against women works in the U.S. Army.

Did you notice the constant feedback between ethnographic and survey data here? The ethnography produced ideas for policy recommendations and for the content of a questionnaire. The questionnaire data illuminated and validated many of the things that the ethnographer learned during participant observation. Those same survey data produced anomalies—things that didn’t quite fit with the ethnographer’s intuition. More ethnography turned up an explanation for the anomalies. And so on. Ethnographic and survey data combined produce more insight than either does alone.

For more on participant observation fieldwork, see Bogdan 1972, Lofland 1976, Spradley 1980, Stocking 1983, Kirk and Miller 1986, Woods 1986, Fine and Sandstrom 1988, Fenno 1990, Burawoy 1991, Behar 1996, Smith and Kornblum 1996, Gummerson 2000, DeWalt and DeWalt 2002, Anderson 2003, and Wolcott 2005.

14



Field Notes: How to Take Them, Code Them, Manage Them

Those who want to use qualitative methods because they seem easier than statistics are in for a rude awakening.

—Taylor and Bogdan 1984:53

Anthropologists collect many kinds of data. We collect audiotapes of musical performances and of recitations of folktales and myths; videotapes of ceremonies, dances, and everyday activities; photographs; newspaper clippings; transcriptions of formal interviews; caches of personal letters. But all anthropologists take field notes. In this chapter, I focus on field notes—how to write them, how to code them, and how to manage them. Many of the lessons about coding and managing field notes apply just as well to transcripts of interviews and to other textual data, which I'll take up in chapter 17.

About Field Notes

Plan to spend 2–3 hours every working day of a participant observation study writing up field notes, working on your diary, and coding interviews and notes. Ralph Bolton asked 34 anthropologists about their field note practices; they reported spending anywhere from 1.5 hours to 7 hours a day on write-up (1984:132).

Remember that it takes twice as long to write up notes *about* a recorded interview as it does to conduct an interview in the first place. You have to listen to a recorded interview at least once before you can write up the essen-

tial notes from it, and then it takes as long again to get the notes down. If you need full transcriptions of interviews, plan to spend 6–8 hours for each hour of interview, assuming that the recording is clear, the interview is in your own language, and you have a transcribing machine with a foot pedal. You can cut transcription time in half by using voice recognition software (more about this back in chapter 9, and see appendix F for information on transcription hardware and software).

Every colleague with whom I've ever discussed this agrees that it's best to set aside a time each day for working on your notes. And don't sleep on your notes. It's easy to forget material that you want in your notes if you don't write them up in the afternoon or evening each day. The same goes for your own thoughts and impressions of events. If you don't write them up every day, while they are fresh, you'll forget them.

This means that you shouldn't get embroiled in a lot of activities that prevent you from writing up field notes. There are plenty of exceptions to this rule. Here's one. You are studying how families create culture by telling and retelling certain stories. You sit down to write up the day's field notes and you get a call from a key informant who tells you to come right on over to meet her father who is leaving on a trip in the morning and wants to tell you himself the story she had told you earlier about his experience as a refugee during World War II. You couldn't possibly turn that one down. But remember, it's easy to let doing anything except writing notes become the norm rather than the exception.

Create many small notes rather than one long, running commentary. Write your notes on a computer and make many separate files—one for each day is fine—rather than adding to the same humongous file day after day. The advantage is that you can name your notes by their date of creation. That way, the computer will present the notes to you in chronological order so you can always find a particular day's (or week's) notes. Many small files are also easier to handle when we get to text management and retrieval programs.

Finally, there are two radically different styles when it comes to writing field notes. Some people like to immerse themselves completely in the local culture and concentrate on the experience. They write up field notes when and as they find the time. Most ethnographers advocate writing up field notes every day, while you are still capable of retrieving detail about the day's events and interactions. I've done both and, like Miles and Huberman (1994), I'm convinced that obsessiveness about writing field notes is the way to go.

How to Write Field Notes

The method I present here for making and coding field notes was developed and tested by the late Michael Kenny and me, between 1967 and 1971, when

we ran those NSF-supported field schools in cultural anthropology that I described in chapter 13. Kenny and I relied initially on our own experience with field notes and we borrowed freely from the experience of many colleagues. The method we developed—involving jottings, a diary, a daily log, and three kinds of formal notes—was used by 40 field-school participants in the United States and in Mexico and by others since then. Some years later, after personal computers came on the scene, my students and I began to think about using machines to help manage textual data (Bernard and Evans 1983).

Two things can be said about the method I'm going to lay out here: (1) It works; and (2) It's not the only way to do things. If you do field research, you'll develop your own style of writing notes and you'll add your own little tricks as you go along. Still, the method described here will help you work systematically at taking field notes and it will allow you to search through them quickly and easily to look for relations in your data. I wish I had used this method when I was doing my own M.A. and Ph.D. fieldwork—and I wish that computers and database management systems had been available then, too.

Four Types of Field Notes

You'll write four kinds of notes in fieldwork: **jottings**, a diary, a log, and field notes proper.

Jottings

Field jottings—what Roger Sanjek calls **scratch notes** (1990:96)—are what get you through the day. Human memory is a very poor recording device, especially for the kind of details that make the difference between good and so-so ethnographic research. Keep a note pad with you at all times and make field jottings on the spot. This applies to both formal and informal interviews in bars and cafés, in homes and on the street.

It also applies to things that just strike you as you are walking along. Jottings will provide you with the trigger you need to recall a lot of details that you don't have time to write down while you're observing events or listening to an informant. Even a few key words will jog your memory later. Remember: *If you don't write it down, it's gone.*

Clearly, there are times when you just can't take notes. Morris Freilich did research in the 1950s with the Mohawks in Brooklyn, New York, and on the Caughnanaga Reservation, 10 miles south of Montreal. He did a lot of participant observation in a bar and, as Freilich tells it, every time he pulled out a notebook his audience became hostile. So, Freilich kept a small notebook in

his hip pocket and would periodically duck into the men's room at the bar to scribble a few jottings (Freilich 1977:159).

William Sturtevant used stubby little pencils to take furtive notes; he found the technique so useful, he published a note about it in the *American Anthropologist* (1959). When Hortense Powdermaker did her research on race relations in Mississippi in 1932, she took surreptitious notes on sermons at African American churches. "My pocketbook was large," she said, "and the notebook in it was small" (1966:175).

Every fieldworker runs into situations where it's impossible to take notes. It is always appropriate to be sensitive to people's feelings, and it is sometimes a good idea to just listen attentively and leave your notebook in your pocket. You'd be surprised, though, how few of these situations there are. Don't talk yourself into not jotting down a few notes on the incorrect assumption that people won't like it if you do.

The key is to take up the role of researcher immediately when you arrive at your field site, whether that site is a peasant village in a developing nation or a corporate office in Chicago. Let people know from the first day you arrive that you are there to study their way of life. Don't try to become an inconspicuous participant rather than what you really are: an observer who wants to participate as much as possible. Participant observation means that you try to *experience* the life of your informants to the extent possible; it doesn't mean that you try to melt into the background and *become* a fully accepted member of a culture other than your own.

It's usually impossible to do that anyway. After four decades of coming and going in Indian villages in Mexico, I still stick out like a sore thumb and have yet to become the slightest bit inconspicuous. Be honest with people and keep your notepad out as much of the time as possible. Ask your informants for their permission to take notes while you are talking with them. If people don't want you to take notes, they'll tell you.

Or they may tell you to take notes when you don't want to. Paul Killworth studied the social organization of the British Army. Since notebooks are, as he says, "part of Army uniform," he was able to go anywhere with his notebook in hand and take notes freely. But if he put his notebook aside for more than a few minutes, soldiers would ask him if he was getting lazy. "More than one relaxing moment," he says, "was stopped by someone demanding that I write something down" (1997:5).

Or they may ask to see your notes. A student researcher in one of our field schools worked in a logging camp in Idaho. He would write up his notes at night from the jottings he took all day. Each morning at 6:00 A.M. he nailed the day's sheaf of notes (along with a pen on a string) to a tree for everyone to look at. Some of the men took the time to scribble helpful (or amusing, or

rude) comments on the notes. If you use this technique, watch out for the CNN effect. That's when people tell you things they want to tell everyone because they know you're going to broadcast whatever they say. This is a disaster if you're trying to make everybody around you feel confident that you're not going to blab about them.

Even when people get accustomed to your constant jottings, you can overdo it. Emerson et al. (1995:23) cite the following field note from an ethnographer who was studying divorce negotiations:

On one occasion when finishing up a debriefing . . . [the mediator] began to apply some eye make-up while I was finishing writing down some observations. She flashed me a mock disgusted look and said, "Are you writing *this* down too!" indicating the activity with her eye pencil.

The Diary

Notes are based on observations that will form the basis of your publications. A diary, on the other hand, is personal. It's a place where you can run and hide when things get tough. You absolutely need a diary in an ethnography project. It will help you deal with loneliness, fear, and other emotions that make fieldwork difficult.

A diary chronicles how you feel and how you perceive your relations with others around you. If you are really angry at someone, you should write about it—in your diary. Jot down emotional highs and lows while they're happening, if you can, and write them up in your diary at the end of the day. Try to spend at least half an hour each day pouring out your soul to a diary. Later on, during data analysis, your diary will become an important professional document. It will give you information that will help you interpret your field notes and will make you aware of your personal biases.

The important thing about a diary is just to have one and to keep it separate from your other field notes. Franz Boas got engaged to Marie Krackowizer in May 1883, just 3 weeks before beginning his first field trip. It was a grueling 15 months on Baffin Island and at sea. Boas missed German society terribly and, although he couldn't mail the letters, he wrote about 500 pages to his fiancée. Here is an excerpt from his extraordinary diary:

December 16, north of Pangnirtung. My dear sweetheart. . . . Do you know how I pass these long evenings? I have a copy of Kant with me, which I am studying, so that I shall not be so completely uneducated when I return. Life here really makes one dull and stupid. . . . I have to blush when I remember that during our meal tonight I thought how good a pudding with plum sauce would taste. But you have no idea what an effect privations and hunger, real hunger, have on a

person. Maybe Mr. Kant is a good antidote! The contrast is almost unbelievable when I remember that a year ago I was in society and observed all the rules of good taste, and tonight I sit in this snow hut with Wilhelm and an Eskimo eating a piece of raw, frozen seal meat which had first to be hacked up with an axe, and greedily gulping my coffee. Is that not as great a contradiction as one can think of? (Cole 1983:29)

February 16. Anarnitung . . . I long for sensible conversation and for someone who really understands me! Unfortunately, this time I did not bring a book to read, so I cannot help myself. I read all the advertisements and everything else on one page of the *Kolnische Zeitung* [a magazine]. In four days I shall have been away eight months. I have heard from none of you for four and a half months. (ibid.:42)

When Malinowski was trapped in the Trobriand Islands during World War I, he too, missed his fiancée and European society and occasionally lashed out at the Trobrianders in his diary (Malinowski 1967:253–254). Fieldwork in another culture is an intense experience, but don't think that you have to be stranded in the Arctic or in Melanesia for things to get intense.

Your diary will give you an outlet for writing things that you don't want to become part of a public record. Publication of Malinowski's and Boas's diaries have helped make all fieldworkers aware that they are not alone in their frailties and self-doubts.

The Log

A log is a running account of how you plan to spend your time, how you actually spend your time, and how much money you spent. A good log is the key to doing systematic fieldwork and to collecting both qualitative and quantitative data on a systematic basis.

A field log should be kept in bound books of blank, lined pages. Some of my students have been able to use handheld computers with schedule-planning software for their logs (Kenneth Sturrock, Lance Gravlee, personal communication), but if you are not completely comfortable with that technology, then stick with a big, clunky logbook, at least 6" × 8" in size so that you can see at a glance what your agenda is.

Each day of fieldwork, whether you're out for a year or a week, should be represented by a double page of the log. The pages on the left should list what you *plan* to do on any given day. The facing pages will recount what you *actually* do each day.

Begin your log on pages 2 and 3. Put the date on the top of the even-numbered page to the left. Then, go through the entire notebook and put the suc-

cessive dates on the even-numbered pages. By doing this in advance, even the days on which you “do nothing,” or are away from your field site, you will have double log pages devoted to them.

The first day or two that you make a log you will use only the right-hand pages where you keep track of where you go, who you see, and what you spend. Some people like to carry their logs around with them. Others prefer to jot down the names of the people they run into or interview, and enter the information into their logs when they write up their notes in the evening. Keep an alphabetized file of 25-word profiles on as many people you meet as you can.

This can be on index cards or on a computer database. The file will make it much easier to remember who you’re dealing with. Before you go into any second or third interview, look up the key biographical information you have about the person (handheld computers are perfect for this sort of thing). During the first couple of minutes of the interview, work in a comment that shows you remember some of those key bio-facts. You’ll be surprised how far that’ll take you.

Jot down the times that you eat and what you eat (especially if you are doing fieldwork in another culture) and write down who you eat with and how much you spend on all meals away from your house. You’d be surprised at how much you learn from this, too.

After a day or two, you will begin to use the left-hand sheets of the log. As you go through any given day, you will think of many things that you want to know but can’t resolve on the spot. Write those things down in your jot book or in your log. When you write up your field notes, think about who you need to interview, or what you need to observe, regarding each of the things you wondered about that day.

Right then and there, open your log and commit yourself to finding each thing out at a particular time on a particular day. If finding something out requires that you talk to a particular person, then put that person’s name in the log, too. If you don’t know the person to talk to, then put down the name of someone whom you think can steer you to the right person.

Suppose you’re studying a school system. It’s April 5th and you are talking to MJR, a fifth-grade teacher. She tells you that since the military government took over, children have to study politics for 2 hours every day and she doesn’t like it. Write a note to yourself in your log to ask mothers of some of the children about this issue and to interview the school principal.

Later on, when you are writing up your notes, you may decide not to interview the principal until after you have accumulated more data about how mothers in the community feel about the new curriculum. On the left-hand page for April 23rd you note: “target date for interview with school princi-

pal.” On the left-hand page of April 10th you note: “make appointment for interview on 23rd with school principal.” For April 6th you note: “need interviews with mothers about new curriculum.”

As soon as it comes to you that you need to know how many kilowatt hours of electricity were burned in a village, or you need to know the difference in price between fish sold off a boat and the same fish sold in the local market, commit yourself *in your log to a specific time* when you will try to resolve the questions. Whether the question you think of requires a formal appointment, or a direct, personal observation, or an informal interview in a bar, write the question down in one of the left-hand pages of your log.

Don't worry if the planned activity log you create for yourself winds up looking nothing like the activities you actually engage in from day to day. Frankly, you'll be lucky to do half the things you think of to do, much less do them when you want to. The important thing is to fill those left-hand pages, as far out into the future as you can, with specific information that you need and specific tasks you need to perform to get that information.

This is not just because you want to use your time effectively, but because the process of building a log forces you to think hard about the questions you really want to answer in your research and the data you really need. You will start any field research project knowing some of the questions you are interested in. But those questions may change; you may add some and drop others—or your entire emphasis may shift.

The right-hand pages of the log are for recording what you actually accomplish each day. As I said, you'll be appalled at first at how little resemblance the left-hand and the right-hand pages have to one another.

Remember, good field notes do not depend on the punctuality of informants or your ability to do all the things you want to do. They depend on your systematic work over a period of time. If some informants do not show up for appointments (and often they won't), you can evaluate whether or not you really need the data you thought you were going to get from them. If you do need the data, then put a note on the left-hand page for that same day, or for the next day, to contact the informant and reschedule the appointment.

If you still have no luck, you may have to decide whether it's worth more of your time to track down a particular person or a particular piece of information. Your log will tell you how much time you've spent on it already and will make the decision easier. There's plenty of time for everything when you think you've got months stretching ahead of you. But you only have a finite amount of time in any fieldwork project to get useful data, and the time goes very quickly.

Field Notes

And now, about field notes. . . . Let's face it: After a hard day trekking all over [town] [the village] [the jungle] [the desert] interviewing people, hanging out, and recording behavior, it's hard to sit down and write up field notes. Sometimes, it's downright intimidating. We know this much about field notes for sure: The faster you write up your observations, the more detail you can get down. More is better. Much more is much better (except, of course, when data are systematically biased, in which case more is decidedly worse).

There are still places where you can't haul in a big desktop computer or find the power to run it, but handheld computers, with real keyboards, now run on small batteries that can be charged from solar power cells. You can write up field notes with a word processor anywhere.

We're not going back to paper field notes, but taking notes on a computer means that you have to be doubly paranoid about protecting those notes. If you have access to the Internet, you can e-mail copies of your notes to a couple of trusted friends or colleagues (encrypt the notes if you need to). The friend or colleague can print a copy of your notes for you, so you've got a hard copy somewhere in case your computer is stolen *and* all your backup disks are destroyed in a fire at the same time. Think this can't happen? Think again. Trees are a renewable resource. Make paper backups. And upload your notes to an Internet server for good measure. If you can't use your college's or university's server (many restrict the file space that students can use), then rent space on a private server (see appendix F).

There are three kinds of field notes: **methodological notes**, descriptive notes, and analytic notes.

Methodological Notes

Methodological notes deal with technique in collecting data. If you work out a better way to keep a log than I've described here, don't just *use* your new technique; write it up in your field notes and publish a paper about your technique so others can benefit from your experience. (See appendix F for a list of professional journals that publish articles on research methods in the social and behavioral sciences.) If you find yourself spending too much time with marginal people in the culture, make a note of it, and discuss how that came to be. You'll discover little tricks of the trade, like the "uh-huh" technique discussed in chapter 9. (Remember that one? It's where you learn how and when to grunt encouragingly to keep an interview going.) Write up notes

about your discoveries. Mark all these notes with a big “M” at the top—M for “method.”

Methodological notes are also about your own growth as an instrument of data collection. Collecting data is always awkward when you begin a field project, but it gets easier as you become more comfortable in a new culture. During this critical period of adjustment, you should intellectualize what you’re learning about doing fieldwork by taking methodological notes.

When I first arrived in Greece in 1960, I was invited to dinner at “around 7 P.M.” When I arrived at around 7:15 (what I thought was a polite 15 minutes late), I was embarrassed to find that my host was still taking a bath. I should have known that he really meant “around 8 P.M.” when he said “around 7.” My methodological note for the occasion simply stated that I should not show up for dinner before 8 P.M. in the future.

Some weeks later, I figured out the general rules for timing of evening activities, including cocktails, dinner, and late-night desserts in the open squares of Athens. Robert Levine has studied the psychology of time by asking people around the world things like “How long would you wait for someone who was late for a lunch appointment?” On average, Brazilians say they’d wait 62 minutes. On average, says Levine, “Americans would need to be back at their office two minutes *before*” the late Brazilian lunch was just getting underway (Levine 1997:136).

When I began fieldwork with the Ñähñu people of central Mexico in 1962, I was offered *pulque* everywhere I went. (*Pulque* is fermented nectar from the maguey cactus.) I tried to refuse politely; I couldn’t stand the stuff. But people were very insistent and seemed offended if I didn’t accept the drink. Things were particularly awkward when I showed up at someone’s house and there were other guests there. Everyone enjoyed *pulque* but me, and most of the time people were too poor to have beer around to offer me.

At that time, I wrote a note that people “felt obliged by custom to offer *pulque* to guests.” I was dead wrong. As I eventually learned, people were testing me to see if I was affiliated with the Summer Institute of Linguistics (SIL), an evangelical missionary group (and, of course, nondrinkers of alcohol) that had its regional headquarters in the area where I was working.

The SIL is comprised of many excellent linguists who produce books and articles on the grammar of the nonwritten languages of the world and translations of the Bible into those languages. There was serious friction between the Indians who had converted to Protestantism and those who remained Catholic. It was important for me to disassociate myself from the SIL, so my methodological note discussed the importance of conspicuously consuming alcohol and tobacco in order to identify myself as an anthropologist and not as a missionary.

Nine years later I wrote:

After all this time, I still don't like *pulque*. I'm sure it's unhealthy to drink out of the gourds that are passed around. I've taken to carrying a couple of six packs of beer in the car and telling people that I just don't like *pulque*, and telling people that I'd be pleased to have them join me in a beer. If they don't offer me beer, I offer it to them. This works just fine, and keeps my reputation of independence from the SIL intact.

Eight years later, in 1979, I read that William Partridge had a similar predicament during his work in Colombia (Kimball and Partridge 1979:55). Everywhere Partridge went, it seems, people offered him beer, even at 7:00 in the morning. He needed an acceptable excuse, he said, to avoid spending all his waking hours getting drunk.

After a few months in the field, Partridge found that telling people "*Estoy tomando una pastilla*" ("I'm taking a pill") did the trick. Locally, the pill referred to in this phrase was used in treating venereal disease. Everyone knew that you didn't drink alcohol while you were taking this pill, and the excuse was perfect for adding a little virility boost to Partridge's reputation. Partridge used his knowledge of local culture to get out of a tough situation.

Methodological notes, then, have to do with the conduct of field inquiry itself. You will want to make methodological notes especially when you do something silly that breaks a cultural norm. If you are feeling particularly sheepish, you might want to write those feelings into your diary where no one else will see what you've written; but you don't want to waste the opportunity to make a straightforward methodological note on such occasions, as well.

Descriptive Notes

Descriptive notes are the meat and potatoes of fieldwork. Most notes are descriptive and are from two sources: watching and listening. Interviews with informants produce acres of notes, especially if you use a recorder and later write down large chunks of what people say. Observations of processes, like feeding children, building a house, making beer, and so on, also produce a lot of notes. Descriptive field notes may contain birth records that you've copied out of a local church registry; or they may consist of summary descriptions of a village plaza, or an urban shopping mall, or any environmental features that you think are important.

The best way to learn to write descriptive field notes is to practice doing it with others who are also trying to learn. Get together with one or more partners and observe a process that's unfamiliar to all of you. It could be a church

service other than one you've seen before, or it could be an occupational process that you've not witnessed. (I remember the first time I saw plasterers hang ceilings: They did it on stilts.)

Whatever you observe, try to capture in field notes the details of the behavior and the environment. Try to get down "what's going on." Then ask informants who are watching the ceremony or process to explain what's going on, and try to get notes down on their explanation. Later, get together with your research partner(s) and discuss your notes with one another. You'll find that two or three people see much more than just one sees. You might also find that you and your partners saw the same things but wrote down different subsets of the same information.

Analytic Notes

You will write up fewer analytic notes than any other kind. This is where you lay out your ideas about how you think the culture you are studying is organized. Analytic notes can be about relatively minor things. When I finally figured out the rules for showing up on time for evening functions in Greece, that was worth an analytic note. And when I understood the rules that governed the naming of children, that was worth an analytic note, too.

As I said in chapter 2, in the section on theory, it took me almost a year to figure out why the casualty rate among Kalymnian sponge divers was going up, while the worldwide demand for natural sponges was going down. When it finally made sense, I sat down and wrote a long, long analytic field note about it. After thinking about the problem for many years, I finally understood why bilingual education in Mexico does not result in the preservation of Indian languages (it's a long story; see Bernard 1992). As the ideas developed, I wrote them up in a series of notes.

Analytic notes are the product of a lot of time and effort and may go on for several pages. They are often the basis for published papers, or for chapters in dissertations and books. They will be the product of your understanding and that will come about through your organizing and working with descriptive and methodological notes over a period of time. Don't expect to write a great many analytic notes, but write them all your life, even (especially) after you are out of the field.

Coding Field Notes

Gene Shelley (1992) studied people who suffer from end-stage kidney disease. Most patients are on hemodialysis. Some are on peritoneal dialysis. The

“hemo” patients go to a dialysis center, several times a week, while the “pero” patients perform a dialysis (called continuous ambulatory peritoneal dialysis, or CAPD) on themselves several times a day.

Figure 14.1 shows two descriptive notes from Shelley’s research. The stuff along the top of each note are codes. First, there’s a delimiter (she used the

<p>\$ 615 8-16-89: 757.3; Dr. H</p> <p>Dr. H explains that in peritoneal dialysis you exchange 2 liters of fluid several times a day (based on body size). Women do it about 3 times and men about 4 times because of larger body size. People mostly do a “dwell” for about 8 hours overnight while they sleep (fluid is inflow into peritoneal cavity and allowed to sit there overnight). Then they do peritoneal dialysis when they wake up and another time or two during the day. Peritoneal dialysis patients are pretty close to being healthy. They have to take medication but you cannot tell them from healthy people, he says.</p>
<p>\$ 742 8-30-89: 57.3, 757.5; Nurse Ralph B.</p> <p>CAPD training takes about a week to 10 days. During this time, the patient comes in every day and receives training. Ralph thinks that when the whole family comes in for the training, the patients do better. They have about 20 CAPD patients right now. Ralph said there are 3 types of CAPD patients: (1) those patients who are already on hemo and in pretty good shape, usually well-motivated. (2) those who are late getting started and are in trouble (medically) and are hurriedly trying to learn the procedure. (It takes 2 weeks to get a catheter inserted and then have it heal. Since this surgery is viewed as “elective surgery,” it can be bumped and rescheduled.) Only after surgery and healing can the training take place. (3) those who have lost a kidney which was transplanted. They are just waiting for another kidney and they view CAPD as temporary and are not that motivated to learn it because they think they won’t be on it long.</p>

Figure 14.1. Field notes from Gene Shelley’s study of kidney disease patients (1992). Reproduced by permission.

dollar sign) that marks the beginning of each note. This lets you pack all the notes together in one big file if you want to but lets a word processor or text management program know where notes begin and end. Next is a unique number that identifies the note in a continuing sequence, starting with 0001. Next is the date.

Then come some numbers that refer to **topical codes** (more on this in the section coming right up). Finally, at the end of the codes at the top of each

field note, there's a cryptic indicator of the person to whom Shelley attributes the information.

When William Partridge interviewed cannabis growers in Colombia, he identified the texts by a letter code and kept the only copy of his notes in a locked trunk (Kimball and Partridge 1979:174). You don't have to be interviewing cannabis growers to be paranoid about keeping your informants' identity secret. You never know what seemingly innocuous information might embarrass or hurt someone if your data fall into the wrong hands.

Topical Codes: The OCM

Shelley used a modified version of the *Outline of Cultural Materials*, or OCM, to code her field notes. The OCM was developed originally by G. P. Murdock in 1950 as a way to index and organize ethnographic materials in the Human Relations Area Files. (The HRAF is a library of ethnographic and archaeological materials on cultures around the world, past and present. More about HRAF and the OCM in the section on content analysis in chapter 17.)

There are 82 big cultural domains in the OCM, in blocks of 10, from 10 to 91. Block 58, for example, covers marriage with codes for nuptials (585), divorce (586), and so on. Other major domains are things like kinship, entertainment, social stratification, war, health, sex, and religious practices. The OCM has gone through several editions over the years. The full version, which you can print out and take with you to the field, is available online at <http://www.yale.edu/hraf/collections.htm> (Murdock et al. 2004).

Every project is unique, so you'll need codes that aren't in the OCM, but you can add decimals (or words) and extend the codes forever. Table 14.1 shows Shelley's adaptation of the OCM code 757 (medical therapy):

TABLE 14.1

Shelley's (1992) Adaptation of the OCM Code 757 on Medical Therapy

757.1	Transplantation
757.2	Hemodialysis
757.3	CAPD (peritoneal dialysis)
757.4	Home dialysis
757.5	Adjustment to dialysis
757.6	Compliance with medical regime
757.7	Machinery involved in dialysis
757.8	Medicines
757.9	Medical test results
757.91	HIV test results

Don't be put off by the lengthiness of the OCM coding list. That is its strength. Lots of anthropologists have used the OCM over the years to code their field notes and other materials. George Foster coded his 50 years of notes on the Mexican community of Tzintzuntán using the OCM, and Robert Van Kemper uses it to code the data that he's collecting on the same community (George Foster, personal communication; Kemper 2002:289). John Honigman used it in his fieldwork on Canadian Indians; the 37 field researchers in Clyde Kluckhohn's Comparative Study of Values in Five Cultures Project all used the OCM to code their notes, as did the fieldworkers on the Cornell University team project studying a village in India (Sanjek 1990:108, 232, 331). You'll only use a fraction of the codes on any given project, but once you start using it in the field, you'll quickly find yourself building supplemental coding schemes to fit your particular needs.

Figure 14.2 shows two more descriptive field notes coded with the OCM. The first is from fieldwork I did in Tarpon Springs, Florida (Bernard 1965) and the second is from a study of an ocean-going research vessel (Bernard and Killworth 1974).

Figure 14.3 shows how Gordon Gibson (of the Smithsonian) used the OCM to code a series of ethnographic films on the Himba, a cattle herding society in Namibia (in Kreiss and Stockton 1980:287). In the first note in figure 14.2, I used the OCM codes for assimilation, cultural goals, vacations, and retirement. I also added a code, K, which marks people's relations to Kalymnos, the island in Greece where they or their parents were born.

In this note, 571 is the OCM code for social relationships and groups. I expanded it to include 571.1, which I call "between-group conflict."

In figure 14.3, the entire film is about a wedding, so each piece is coded 585, the OCM code for nuptials. Where the hut is seen, the code for dwellings (342) is inserted. Where the film shows people eating meat, the codes 262 (diet) and 264 (eating) are inserted. In the frame at 11:45:10, the code for visiting (574) appears, and in the two earlier frames, the code for childhood activities (857) is inserted. When Gibson did this work in the 1970s, the database was held on a mainframe. Today, you can watch an ethnographic film on DVD and enter codes into a database as you go. (For more on this, see the section on database management, below, and chapter 17 on text analysis).

Topical Codes: Other Coding Schemes

There are, of course, other useful checklists for coding cultural data. Allen Johnson and his colleagues developed one in conjunction with Johnson's Time Allocation Project (see chapter 15 on direct observation methods). Like the OCM and like *Notes and Queries in Anthropology*, the Johnson et al. (1987) checklist can be used as a reminder of the kinds of data you might want to

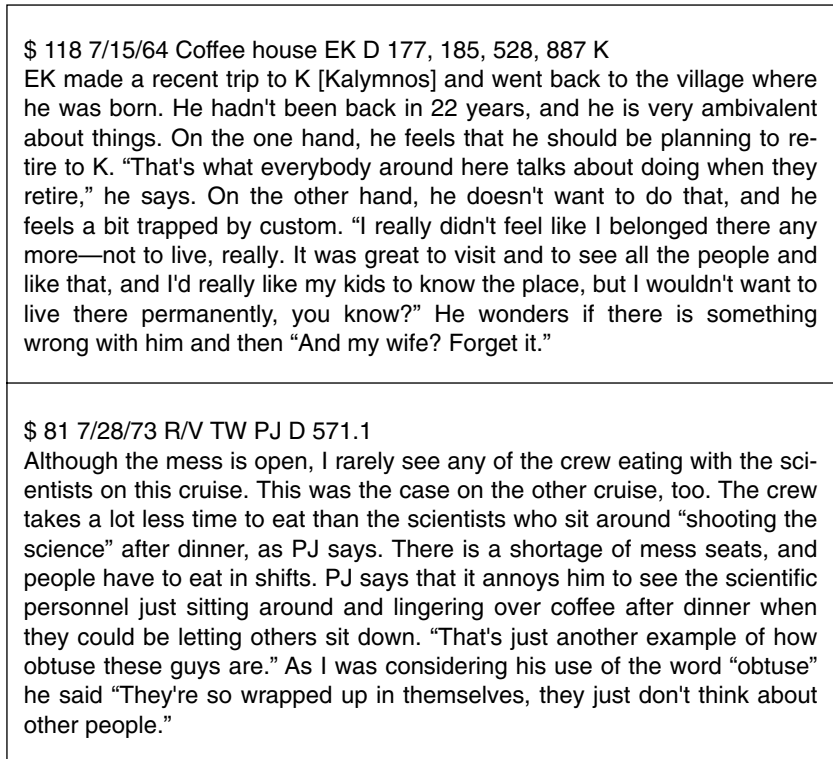


Figure 14.2. Two field notes.

collect in a general ethnographic research project and as a template for coding data collected during field research.

Many people find the use of number codes distracting. Matthew Miles and Michael Huberman (1994), authors of a wonderful book on qualitative data analysis, advocated the use of words or mnemonics that look like the original concept. Like many researchers, they find that mnemonic codes (like ECO for economics, DIV for divorce, and so on) are easier to remember than numbers. Figure 14.4 shows a piece of Miles and Huberman's codebook for a study they did of innovations in a school system.

Another value of using your own codes is that they develop naturally from your study and you'll find it easy to remember them as you code your notes each day. Strauss and Corbin (1990:68) recommend *in vivo* codes as names for things. *In vivo* codes are catchy phrases or words used by informants. In

13 GIBSON (film)		E-5 (1961) 1969	HIMBA
11:09:29	Picture	Bride and companion walk toward hut, then bride and unmarried girl drop to their knees and crawl into hut. People are seen sitting in front of hut as the two disappear inside.	585* 342
	Sound	The bride and their companions solemnly return to the hut in Vesenga's village where she has been staying, and she and the unmarried girl enter the hut.	
11:29:30	Picture	People sitting and standing near hut.	585*
	Sound	Women and children of the village, and those visiting, from other villages, have gathered to sit near the bride.	342 574 857
11:42:22	Picture	Boys seated eating meat.	585*
	Sound	Young boys eat together.	262 857
11:45:10	Picture	Meat in basket and men seated, eating. A man standing and eating meat off a bone, places the bone on a bush. The groom is seated, his arms folded on his knees. He takes a piece of meat from a pail on the ground between his feet.	585* 264 574
	Sound	The bridegroom and his friends are still seated by the bower, where they are finishing.	

Figure 14.3. Gibson's Coding of the Himba Films.

SOURCE: L. Kreiss and E. Stockton, "Using the Outline of Cultural Materials as a Basis for Indexing the Content of Ethnographic Films," *Behavior Science Research*, Vol. 15, pp. 281-93, 1980.

<i>Short Description</i>	<i>Code</i>
Adoption Process	AP
AP:Event chronology-official version	AP-chron/pub
AP:Event chronology-subterranean	AP-chron/prov
AP:Inside/outside	AP-in/out
AP:Centrality	AP-cent
AP:Motives	AP-mot
AP:User fit	AP-fit
AP:Plan	AP-Plan
AP:Readiness	AP-Redi
AP:Critical events	AP-crit

Figure 14.4. Part of Miles and Huberman's coding scheme.

SOURCE: M. B. Miles and A. M. Huberman, *Qualitative Data Analysis*, p. 59. © 1994 by Sage Publications. Reprinted by permission of Sage Publications.

his study of Alaskan fishermen, Jeffrey Johnson heard people talking about a “clown.” The word turned out to be a terrific label for a type of person found in many organizations. The term emerged *in vivo* from the mouths of Johnson’s informants. (For more on *in vivo* coding, see the section in chapter 17 on grounded theory in text analysis).

A good way to find *in vivo* codes is to use a computer program to count the number of times that words occur in your text (see appendix F for programs that do this). The nontrivial words that appear often are hints about themes in your notes. When you have a list of codes words, use your word processor to grab the codes and put them into your notes wherever you want.

If you use your own coding scheme, or if you modify an existing scheme (like the OCM), be sure to write up a verbose codebook in case you forget what “A5” or “EMP” or whatever-cute-abbreviations-you-dreamed-up-at-the-time-you-did-the-coding mean.

And don’t get too picky when you make up your own codes. Coding is supposed to be data reduction, not data proliferation. Matthew Miles was involved in a big ethnographic project to evaluate six schools. All the researchers developed their own codes and the code list quickly grew to 202 categories of actors, processes, organizational forms, and efforts. Each of the six researchers insisted that his or her field site was unique and that the highly specialized codes were all necessary. It became impossible for anyone to use the unwieldy system, and they just stopped coding altogether (Miles 1983:123).

The important thing is not which coding scheme you use, it’s that you code your notes and do it consistently. In most projects, the coding scheme takes shape as the notes are written. The scheme is revised a lot before it becomes stable. Some anthropologists, even those who use the OCM, wait a month or more, to see how their field notes are shaping up, before they think about how to code the notes.

The Mechanics of Coding

Put codes directly into the text of the field notes, either at the top of each note, or to the right or left side of each note. Figure 14.5 shows an example of coding across the top and of coding down the edge, using mnemonic codes.

The choice of top coding or edge coding is really a matter of taste—you can do one as easily as the other in your word processor. If you choose edge coding, just define a narrow column of, say, 15 or 20 spaces for the codes and a broad column of, say, 50 or 55 columns for the body of text.

Coding across the top	
412 MA XOR 101290 MIG WOM ECO This note is number 412. It's about an informant named MA in these notes, and she is from a village you label XOR. The date is October 12, 1990, and the note is about MA's migration from her village to the city in search of work. The note is coded using abbreviations as being about migration (MIG), about women (WOM), and about economics (ECO).	
Coding down the edge	
412, MA XOR 101290 MIG WOM ECO	This note is number 412. It's about an informant named MA in these notes, and she is from a village you label XOR. The date is October 12, 1990, and the note is about MA's migration from her village to the city in search of work. The note is coded in abbreviations as being about migration (MIG), about women (WOM), and about economics (ECO).

Figure 14.5. Coding field notes.

Coding vs. Indexing

I want to make clear the three different uses of the word “code.” When I say: “Use codes for places and informant names,” the word “code” means an *encryption device*. The object is to hide information, not dispense it. When I say: “After the informant name and place, leave room for topical codes,” the word “code” means an *indexing device*. The object is to identify the existence of a variable. The third meaning of the word “code” is a *measurement device*.

Suppose you do 100 interviews with women about their birthing experience. If you stick the code PAIN into the text whenever a woman mentions anything about pain or about feeling hurt, you’d be using the code PAIN as an indexing device—that is, as a way to find your way back to all the places in the text where anything about pain is mentioned. It’s just like the subject index to this book. It says that “quota sampling” is on page 187 and sure enough, when you go to page 187, you find you’re reading about sampling.

Suppose, though, that you make judgments about the amount of pain—by counting words like “agony” as indicating more pain than words like “distress” or by looking at the content and meaning of the text and counting “It was painful, but I got through it” as indicating less pain than “I prayed I would die.” You might use codes like LO-PAIN, MID-PAIN, or HI-PAIN and in this case, you’d be using codes for actual measurement. We’ll talk more about this in chapter 17 on text analysis.

Analyzing Field Notes

I think it's best to start with the ocular scan method, or eyeballing. In this low-tech method, you quite literally lay out your notes in piles on the floor. You live with them, handle them, read them over and over again, tack bunches of them to a bulletin board, and eventually get a feel for what's in them. This is followed by the interocular percussion test, in which patterns jump out and hit you between the eyes. (If you take your notes on a computer, print them for this part of the analysis.)

This may not seem like a very scientific way of doing things, but I don't know any way that's better. No researcher, working alone for a year, can produce more field notes than she or he can grasp by pawing and shuffling through them. For sheer fun and analytic efficiency, nothing beats pawing and shuffling through your notes and thinking about them.

Eventually, though, you'll want to use a text management (TM) program. It's not unusual for anthropologists to produce 10,000 words some weeks in field notes. That's the equivalent of a 40-page, double-spaced paper, or about six pages a day. This won't happen every week, but when you're doing fieldwork, bombarded all day long with new sensory experiences and interviewing informants about topics that really mean something to you, writing 40 pages of field notes is easy. Even at a more modest clip, you're likely to accumulate 500–1,000 pages of notes in a year.

That pile of field notes can get pretty intimidating, and the next thing you know you're backing away from taking a lot of notes on the theory that fewer notes are easier to handle. I know; it happened to me, and it has happened to many of our colleagues. TM software can help.

TM programs make light work out of searching through tons of notes for combinations of words or phrases. With a TM program, you can ask the computer questions like: "Find every note in which I used the word *woman* but only if I also used the word *migration* within three lines of the word *woman*." Since codes are just words, you can substitute OCM code numbers or in vivo codes or mnemonic codes in the queries. For example: If you've left up to three lines at the top of each note for topical codes, you'd ask the computer: "Find 166 and 572.1 if they occur within three lines of one another."

TM programs may also tempt you to avoid coding your field notes. Resist this temptation. You can watch a wedding ceremony for 3 hours, spend a day and a half writing up 22 pages of notes on your observations, and never use the words "marriage" in your notes. If you use a text manager to find the word "marriage," you'll retrieve all the notes in which you *did* use that word, but you won't find any of the 22 pages where you wrote about that wedding ceremony you attended.

If you do edge coding, you need a slightly different search strategy. Suppose you have 33 lines of text on each double-spaced page of field notes or interview transcription. You could have codes anywhere along the edge of the page from line 1 to line 66. Decide on a marker, like the dollar sign, to delimit the top and bottom of each field note. Then ask a TM program to look for a pair or a series of codes between consecutive markers. Figure 14.6 shows one of Gene Shelley’s notes coded along the edge.

<p>\$</p> <p>9589Inf#2</p> <p>57</p> <p>572</p> <p>752.2</p> <p>76</p> <p>157</p> <p>\$</p>	<p>\$</p> <p>She says the social network is important, in her opinion. It is important to have a person to support you (emotionally and physically). Her friends and school children (she was a teacher for deaf students for a short time), called her to see how she was doing and they still get in touch with her now. However, her friends cry and say what a brave person she is and she doesn't like them to cry. "People get upset, so I don't talk to them about it." She also doesn't like to talk to people for another reason. She started dialysis in 1972. Since then, all others who started with her (in her cohort) are dead. She doesn't want to meet new people. She doesn't want to talk to other patients about personal stuff because she will get attached to them and they will die (or suffer horribly with another disease like diabetes). Even with people who are not sick, she doesn't always tell everyone about CAPD. She would rather talk to them about normal things, not her disease.</p> <p>\$</p>
---	--

Figure 14.6. Gene Shelley’s edge-coded field note.

Gregory Truex (1993) marks off chunks of text using what he calls the “B–E convention.” He marks the beginning of a chunk of related text with a code in angle brackets and tacks on the letter “B,” for beginning. Then he uses the same code, again in angle brackets, with the letter “E” tacked on to mark the end of the chunk.

For example, Truex marks the beginning of any text about land use with <LANDB> and marks the end of the block with <LANDE>. Then he uses a TM program to retrieve chunks of text that begin with <LANDB> and end

with <LANDE> or chunks that begin with <LANDB>, end with <LANDE>, and contain some third code, like <AGRIC>. Truex finds that this process of “tagging and tying,” as he calls it, helps him understand his corpus of text.

Database Management and Paper Notes

But what do you do if you have a big pile of paper notes from projects gone by? Or a stack of local newspaper clippings? The solution to this problem is database management, or DBM.

We use DBM all the time. If someone asks you to suggest a French restaurant in San Francisco that costs less than \$100 per person (with wine), you search through your mental database of restaurants, limiting yourself to those you know in San Francisco (the first criterion). Then you look through *that* list for French restaurants and pull out only those that also satisfy the second criterion. Finally, you look through *that* list and see if any of them cost less than \$100 per person (the third criterion).

If you have paper copies of interview transcripts or field notes and no computer file, number the notes, starting with 00001. Set up a database using Microsoft Access® or FileMaker Pro®. The records of the database will be the numbered pages, from 1 to *n*, and the fields of the database will include the name of the informant, the place where you did the interview, the date, and the topics covered in each note. Some notes may get one or two topical codes; others may need 10, so build 10 code spaces into the database.

It takes about 2 or 3 minutes per note to enter codes into a database, which means it only takes 30–50 hours of work at the computer to enter codes for 1,000 pages of field notes. When you ask the database “Which notes are about religion and also about political factionalism but not about generational conflict?” you’ll get back answers like: “The information you want is on records (that is, pages) 113, 334, 376, 819, 820, and 1,168.” You simply flip through the field notes on your lap and as you do, you’ll see the entire page of each field note and you’ll get a feel for the context.

You can use DBM software to manage information about paper notes, photos and slides, collections of artifacts, collections of news clippings—in short, any collection of *things*. You could code each CD you own for artist, title, publisher, date of release, and topics that describe the music. If you have 1,000 CDs, then you have 1,000 records in the database. You could code all the photos you take in the field and make each one a record in a database. Then you can ask questions like: “Which photos are about palm oil?” or “Which photos are about old men in the plaza?” or “Which are about market sellers *and* about meat *and* about servants making purchases for others?” There is nothing more

qualitative than things, so this is what qualitative database management is really about.

Local newspapers are a great source of data about communities, but once you start clipping all the interesting stories, you quickly wind up with hundreds of pieces. Just number them and code them for the topics that are germane to your research. Then, later, you can ask the database: “Which clippings are about property disputes between siblings?”

Relational Database Systems

All the database examples I’ve given you so far are for **flat files**, which means that the data are in a single table. Suppose you do 50 semistructured interviews. A 2-hour interview typically produces about 30 pages of transcribed, double-spaced text. If you code, on average, five chunks of text on each page, you’ll wind up with

$$50 \text{ (informants)} \times 30 \text{ (pages)} \times 5 \text{ (chunks)} = 7,500$$

records in a flat file database—one record for each chunk.

Now, suppose that one of the things you code for is whether a particular chunk of text is about the informant’s experience with migration. And suppose you want to ask the database: “Which pieces of the texts are about migration and about women who are married and under 30 years of age?” Table 14.2

TABLE 14.2
A Flat File Database

<i>Person</i>	<i>Sex</i>	<i>Age</i>	<i>Chunk</i>	<i>Married</i>	<i>Code 1</i>			
					<i>Migration</i>	<i>Code 2</i>	<i>Code 3</i>	<i>Code 4</i>
Josefa R.	2	22	1	Y	Y			
Josefa R.	2	22	2	Y	N			
Josefa R.	2	22	3	Y	N			
Josefa R.	2	22	.	Y	N			
Josefa R.	2	22	.	Y	Y			
Josefa R.	2	22	148	Y	N			
Roberto M.	1	47	1	N	N			
Roberto M.	1	47	2	N	N			
Roberto M.	1	47	3	N	N			
Roberto M.	1	47	.	N	N			
Roberto M.	1	47	.	N	N			
Roberto M.	1	47	160	N	N			

shows a microscopic piece of the flat file database you'd need to answer this question. Each informant has a set of records, one record for each chunk of text. You can see the problem: Each record contains information about the informant's age, sex, and marital status, plus information about the themes (codes) that are present on each page of text.

There's a better way. In a **relational database**, you'd have one file with information about the informants (you'd type that information into the database once for each informant—that is, 50 times, not 7,500 times) and another file that contains information about their transcripts.

Each of the 7,500 records in the transcript data file would contain a single code that identifies the informant and the two files would be related to each other through that code. So when you ask: "Which pieces of text are about migration and about women under 30?" the program: (1) rushes off to find all the chunks of text that you've tagged for migration; (2) makes a temporary list of those chunks; (3) looks at each chunk in the temporary list and finds the respondent's name or code; and (4) looks up the respondent in the respondent information file.

If the respondent is not a woman, or is not a woman under 30, the program drops the chunk of text from the temporary list. When the program finishes, it tells you whether any of the chunks of text conform to all the criteria you listed in your question. Asking questions like this for 7,500 records takes a couple of seconds.

Here's an example of a really interesting DBM problem in anthropology. Among the Embu of Kenya, women between the ages of 15 and 49 spend about 80% of their lives either pregnant or lactating. They average 30 months between births and, on average, breast-feed their children for 15 months. In 1985–1986, Michael Baksh and his colleagues did a 12-month time allocation study of 169 Embu households comprising 1,318 persons (Baksh et al. 1994:346). The researchers made 64 random visits to each household. They recorded what everyone was doing at the moment of the visit and they coded each lead woman (one per household) as being in one of nine reproductive status periods: the 1st, 2nd, and 3rd trimesters of pregnancy; the 1st, 2nd, 3rd, 4th, and 5th 3-month period of lactation; or NPNL (nonpregnant, nonlactating).

In the late 1980s, data had to be stored on magnetic tapes and analysis was done on a mainframe computer. Today, Baksh and his colleagues could use relational database software that operates on personal computers and set up four linked files. In one file would be data about the households (number of people, the material the house was made of, and so on). This is shown in figure 14.7a. In another file would be information about the 1,318 people in the 169 households (sex, date of birth, etc.). This is shown in figure 14.7b.

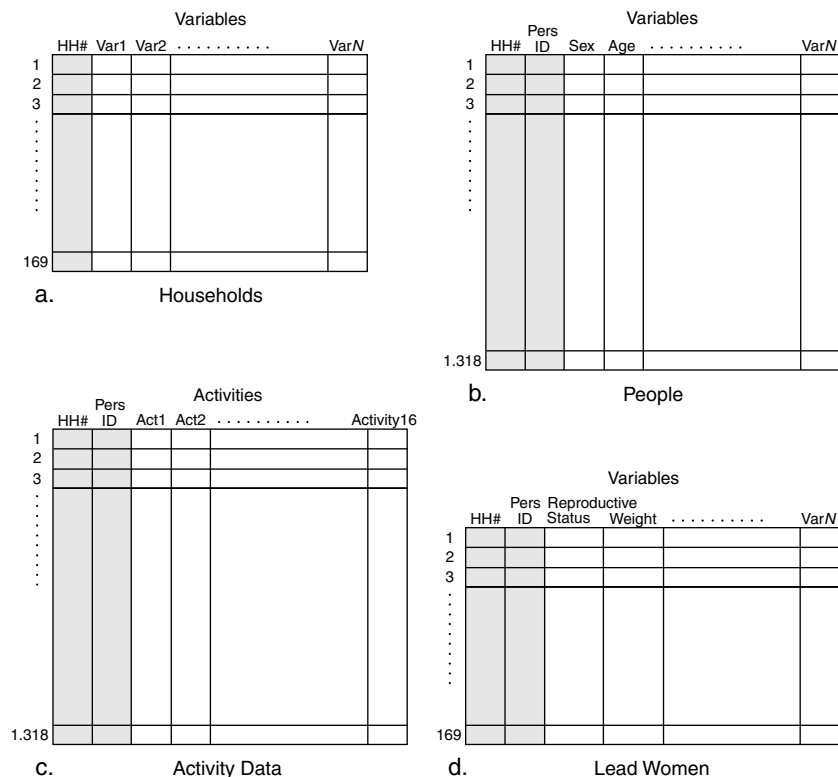


Figure 14.7. Schema for Baksh’s relational database.

Baksh et al. coded for hundreds of activities, but these were combined into 16 categories for analysis (eating, subsistence agriculture, food preparation, child care, and so on). These data would be stored in a summary table of activities for the 1,318 people, as shown in figure 14.7c. Note that since there were 64 visits, the numbers in each of the 16 activity columns would range from 0 to 64.

Data on the monthly reproductive status of each lead woman would be stored in a set of 12 files, along with information about her physical functions, body mass index, and so on. This is shown in figure 14.7d.

Do you see how the files in figure 14.7 are related to one another through common features? File (a) is linked to files (b), (c), and (d) through the unique household identifier (the variable labeled HH#). Thus, for any person in file (b), we have access to all the data about the household they belong to through

this link. File (b) is linked to files (c) and (d) through the unique personal ID, or identification number. So, for each of the 169 lead women in file (d) we have access to her personal information (age, etc.) stored in file (b). The information in file (c) is linked to information about individuals and about households through the personal and household ID numbers.

Once you get accustomed to setting up databases, you'll wonder how you ever got along without them.

15



Direct and Indirect Observation

You can observe a lot by just watching.

—Yogi Berra 1964, cited in Berra and Garagiola 1998

Interviewing is a great way to learn about attitudes and values. And it's a great way to find out what people think they do. When you want to know what people *actually do*, however, there is no substitute for watching them or studying the physical traces their behavior leaves behind. This chapter is about **direct observation** (watching people and recording their behavior on the spot) and **indirect observation** (the archeology of human behavior).

There are two big strategies for direct observation of behavior. You can be blatant about it and **reactive**, or you can be unobtrusive and **nonreactive**. In reactive observation, people know that you are watching them and may play to their audience—you. You can wind up with data about what people want you to see and learn little about what people do when you're not around. In unobtrusive observation, you study people's behavior *without their knowing it*. This stops people from playing to an audience, but it raises tough ethical questions. We'll get to some of those problems later in this chapter.

We begin with the two most important methods for direct observation, **continuous monitoring** and **spot sampling** of behavior. Then we take up **unobtrusive observation** (and the ethical issues associated it), and, finally, indirect observation (trace studies and archival research).

CM—Continuous Monitoring

In continuous monitoring, or CM, you watch a person, or group of people, and record their behavior as faithfully as possible. The technique was devel-

oped in the field of management by Charles Babbage, the 19th-century mathematician who invented the computer. He studied the behavior of workers in a factory and determined that a pound of number 11 straight pins (5,546 of them) should take exactly 7.6892 hours to make (Niebel 1982:4).

CM is still used in assessing work situations (Chadsey-Rusch and Gonzalez 1988; Drury 1990; Frank et al. 1997), employee-employer interactions (Sproull 1981), teacher-pupil interactions (LeCompte 1978; Guilmet 1979; LeCompte et al. 1993; Meh 1996), police-civilian interactions (Black and Reiss 1967; Reiss 1971; McCall 1978; Sykes and Brent 1983), and the behavior of long-term patients in hospitals and nursing homes (Algase et al. 1997).

CM is the core method of **ethology** (Hutt and Hutt 1970; Sullivan 1990; Houck and Drickamer 1996; Lehner 1996; Zinner et al. 1997). Most ethologists study nonhuman animals (everything from moths to fish to chimpanzees), but Darwin (1998 [1872]) used direct observation of facial expressions to study emotions in humans and animals—an area of interest ever since (Eckman 1973, 1980). CM is a mainstay in behavioral psychology for assessing anxieties and phobias (Dadds et al. 1994), and it has been used to study how people eat (Stunkard and Kaplan 1977) and how people use architectural space (Bechtel 1977). CM is one of the all-around varsity methods and there is increasing interest in using ethological methods for the study of natural human behavior. (See Eibl-Eibesfeldt [1989] for the definitive resource on human ethology and see also Atzwanger et al. [1997]. See Chisholm [1983] for an ethological study of Navajo children.)

Ethograms

It is standard practice in ethology to develop an **ethogram**, or list of behaviors, for a species being studied. It's painstaking work. Kneidinger et al. (2001) studied touching behavior in 119 mostly white, male baseball players and 52 mostly white, female softball players in six major universities in the southeastern United States. Kneidinger et al. developed an ethogram of 37 touching behaviors before they even launched their main study. The touching behaviors included things like tapping gloves, high fives, butt slaps, and chest grabs ("one participant grabs the front of the other participant's shirt").

The main study involved watching and recording 1,961 touching behaviors across 99 innings of baseball for the men and 1,593 touching behaviors across 63 innings of softball for the women. Among the interesting results: men and women touched each other the same amount after winning games, but women touched each other more than men touched each other after losing games (Kneidinger et al. 2001:52).

Continuous Monitoring in Anthropology

CM has a long and noble history in anthropology. When Eliot Chapple was still a student at Harvard in the 1930s, he built a device he called the “interaction chronograph” for recording on a rolling sheet of paper the minute features (facial expressions, gestures) of human interaction. The interaction chronograph is, as far as I can tell, the unheralded forerunner of the handheld computer recording systems used for continuous monitoring in ethology, psychology, and anthropology today (Chapple 1940; Chapple and Donald 1947).

In 1949, John Roberts and a Zuni interpreter took turns sitting in one of the rooms of a Zuni house, simply dictating their observations into a tape recorder. (That recorder, by the way, was the size of a suitcase and weighed 30 pounds.) This went on for 5 days and produced data for a 75,000-word book, rich in detail about everyday Zuni life. Figure 15.1 shows some excerpts from Roberts’s work.

People let Roberts park in their homes for 5 days because Roberts was a participant observer of Zuni life and had gained his informants’ confidence. Even earlier, in 1936–1937, Jules and Zunia Henry did fieldwork among the Pilagá Indians of Argentina. Among the data they collected was a set of direct observations of children. Table 15.1 shows the data from observations made on four children for 10 kinds of behaviors associated with eating and food sharing.

The data in table 15.1 were extracted from 843 observations of children’s behavior. Here are two of those observations from the original data:

The three children of Diwa’i are feeding peacefully together. Deniki, the baby, waves his hand for food and mother gives him a small piece of palm dipped in fat. After eating a second piece he is given the breast.

Deniki, Naicho, and Soroi are together. Deniki is holding a dish with a very small quantity of cooked fruit in it. Soroi says, “Share it with me,” and takes one fruit out of the dish. Naicho immediately snatches another one away violently, but not before Deniki has already taken one out, which he then offers to Naicho, appearing not to comprehend her action. (Mensch and Henry 1953:467)

The Zapotec Children Study

Douglas Fry used CM to study aggressive play among Zapotec children. From 1981 to 1983, Fry did 18 months of participant observation fieldwork in La Paz and San Andrés, two small Zapotec-speaking villages just 4 miles apart in the Valley of Oaxaca, Mexico. During the last 5 months of his research, Fry did direct, continuous monitoring of 24 children (3–8 years old) in each village. Before that, he visited almost all the households in the villages

0940

E1DaE1S09 is dressed in blue denim overalls and blue denim shirt. FaSiSoSo is wearing a cotton shirt, heavy trousers, jacket and oxfords. The girls are wearing dresses, socks, and shoes. 2Da24 has on a blouse, skirt, green socks, and oxfords.

0941-(FaSiSo37D)

YoDaSo1 came into SCR from ESCR carrying a little toy in his hand.

0945-(FaSiSo37D)

I intended going to the buck herd today to take out my bucks (rams). I was going to bring them down to Zuni to feed them to get them in good shape—but there is no time to go over there today. I think I will go tomorrow.

AdE1So27A went into ESCR, ENCR, and SER, but he had nothing to report. Mo61 is still in SER shelling corn.

0950

Mo61 walks back into WNCR to build a fire in the WNCR cooking stove.

AdE1So27A says that she is going to make hominy with the stew.

3Da22 is mounting turquoise on sticks for grinding.

YoDaSo1 came into SCR a few moments ago with a homemade cardboard horse which had been cut out by YoDaHu22.

2Da2Da3 followed YoDaSo1.

This house is full of activity and the children are running back and forth.

They are not playing outside today because the weather is poor.

E1Da28 is mounting turquoise on sticks in preparation for grinding. She has a fire going in WR, which is a very large room to heat.

Figure 15.1. Excerpts from Roberts's observations of a Zuni household. Persons and things are identified by shorthand notation. For example, 2Da2Da3 is the family's second daughter, who is 3 years old. The sequence begins at 9:40 A.M. and ends at 10:00 A.M.

SOURCE: J. M. Roberts, *Zuni Daily Life*. © 1965 [orig. 1956], HRAF Press.

several times so that children had become accustomed to him when he began his intensive observation.

Fry describes his data collection procedures clearly:

The formal focal sampling observations were conducted between May and September of 1983. They represent each day of the week and encompass the daylight

TABLE 15.1
Nutritional Behavior of Four Pilagá Indian Children

<i>Behavior</i>	<i>Yorodaikolik</i> (4 years)	<i>Tanpani</i> (8–9 years)	<i>Naicho</i> (6 years)	<i>Deniki</i> (15 months)
Given food	9	7	27	42
Deprived of food	7	0	2	4
Deprivation followed by restitution	1	0	1	2
Attempt made to deprive of food	2	1	1	4
Gives food	8	7	1	3
Withholds food	6	5	2	3
Deprives others of food	2	6	1	0
Attempts to deprive others of food	1	5	0	1
Receives part of a whole	0	1	0	4
Punished while eating	0	0	0	3
Total observations of each child from which these are taken	190	207	238	208

SOURCE: I. N. Mensh and J. Henry. 1953. Direct Observation and Psychological Tests in Anthropological Field Work. *American Anthropologist* 55(4): 466.

hours. Most observations (84%) were conducted within family compounds, although children were also observed in the streets, town squares, school yards, fields, and hills. I alternated sampling between the two communities on a weekly to biweekly basis. A total of 588 observations were conducted, resulting in an average of approximately 12 observations for each focal child ($M = 12.25$, $SD = 6.21$). On average, each focal child was observed for just over 3 hours ($M = 3.13$ hours, $SD = 1.39$ hours), resulting in a total of 150 hours of observation time for the entire sample. [It is common in scientific papers to report means and standard deviations; hence the M and SD figures in this paragraph. HRB]

Focal observations were narrated into a tape recorder carried in a small backpack or recorded on paper using a shorthand system. I recorded a running commentary of the behaviors engaged in by the focal child, using behavior elements defined in the previously developed ethogram. I moved with a focal child in order to maintain continuous visual contact (Altmann 1974), but did not remain so close as to interfere with actions or unduly attract the child's attention. Whenever a focal child engaged in any type of antagonistic behavior, the specifics of the interaction were noted, including identity of the interactant(s) and any facial expressions or gestures. For instance, interactions such as the following were recorded: Focal boy punches, pushes sister of 3 year old while laughing (sister does nothing in response). (Fry 1990:326–327)

Fry developed his ethogram of Zapotec children by watching them in public places before beginning his study of focal individuals. Based on 150 hours of

focal child observation, Fry's data contain 764 episodes of what he calls "play aggression" and 85 episodes of "serious aggression."

Play aggression is a punch, kick, tackle, etc., accompanied by smiles, laughs, and playfaces. Serious aggression are episodes accompanied by low frowns, bared teeth, fixated gazes, and crying. Fry found that when girls initiated serious aggression, it was almost always with other girls (93% of cases). But when boys initiated serious aggression, it was just as likely to be with girls as with other boys.

For more on methods of studying children's behavior under natural conditions, see Fine and Sandstrom (1988) and Pellegrini (1996). See Longabaugh (1980) for a review of the uses of direct observation in cross-cultural psychology.

The Samoan Energy Study

Pearson (1990) used CM to study how Samoans in Western Samoa, American Samoa, and Honolulu used energy. Pearson had the idea that, as Samoans moved to Honolulu and became more urbanized, there would be changes in lifestyle and that these changes would show up in their energy intake and expenditure. Pearson asked people to recall their activities over the past 24 hours, and then, to check the **24-hour recall** data, he and a female assistant monitored 47 men and 43 women.

The team did 825 hours of observation and had their subjects in direct view 92% of the time. The estimates of energy expenditure from direct observation of the 47 men were 33%–80% lower than the estimates from the recall data. For women, the estimates were 27%–53% lower. Women had better recall of their activities, but men and women were both way off the mark, particularly in recalling their light-to-moderate work of the previous day. Pearson's work makes it clear that, when it comes to measuring energy expenditure, recall is not a good substitute for observation.

The Shopping Study

Martin Murtagh (1985) used CM to study the use of arithmetic by grocery shoppers. He recruited 24 adults in Orange County, California. Accompanied by two observers, each informant wore a tape recorder while shopping at a supermarket. As the informants went about their shopping, they talked into the tape recorder about how they were deciding which product to buy, what size to choose, and so on.

One observer mapped the shopper's route through the store and recorded the prices and amounts of everything purchased. The other researcher kept up

a running interview with the shopper, probing for details. Murtagh was aware of the potential for reactivity in his study. But he was interested in understanding the way people thought through ordinary, everyday arithmetic problems, and his experiment was a good way to generate those problems under natural conditions.

Many CM researchers record their own observations. It's less tedious than writing; it lets you focus your eyes on what's going on; it lets you record details later that might be left out of an on-the-spot written description; it avoids the limitations of a checklist; and it lets you get information about context as well as about the behavior you're studying. Moreover, you can easily transcribe your recorded observations, once you've got your voice recognition software trained (see above, chapter 9, and appendix F).

But there are trade-offs. If you want measurements from qualitative data (like running commentaries on tape), you have to code them. That is, you have to listen to the tapes, over and over again, and decide what behaviors to code for each of the people you observe. Coding on the spot (by using a behavioral checklist or by inputting codes into a handheld computer) produces immediate quantitative data. You can't code *and* talk into a recorder at the same time, so you need to decide what kind of data you need and why you need them before you choose a method.

If you are trying to understand a behavioral *process*, then focus on qualitative data. If you need measurements of *how much* or *how often* people engage in this or that behavior, then focus on quantitative data. And, as always, who says you can't do both?

Coding Continuous Monitoring Data

Go to a shopping mall and record the interaction behavior of 30 mother-child pairs for 5 minutes each. Record carefully the number of children each mother has and her interaction with each child. Try to find out whether interaction patterns are predictable from: (1) the number of children a mother has to cope with; (2) the ages of the children; (3) the socioeconomic class or ethnicity of the family; or (4) some other factors.

This exercise is instructive, if not humbling. It's a real challenge to code for socioeconomic class and ethnicity when you can't talk to the people you observe. Do this with at least one colleague so you can both check the reliability of your coding.

In hypothesis-testing research, where you already know a lot about the people you are studying, you go out to observe armed with a coding scheme worked out in advance. The idea is to record any instances of behavior that

conform to the items in the scheme. This allows you to see if your hunches are correct about conditions under which certain behaviors occur. In some studies, you might be interested in noting instances of aggressive vs. submissive behavior. In other cases, those variables might be irrelevant.

Coding Schemes

Just as with attitude scales and surveys (in chapters 10 and 12), there's no point in reinventing the wheel. Over the years, researchers have developed coding schemes for using direct observation in many different situations—in studies of interactions between married couples, in studies of teacher effectiveness, in worker-management negotiations, and so on. If others have developed and tested a good system for coding behaviors of interest to you, use it. Don't feel that it's somehow more prestigious or morally better for you to make up everything from scratch. Knowledge grows when researchers can compare their data to the data others have collected using the same or similar instruments.

Figure 15.2 shows the basic coding scheme for **Interaction Process Analysis**, a system developed over half a century ago by Robert F. Bales in his research on communications in small groups (Bales 1950, 2001).

Despite its age, the Bales coding scheme continues to be used in the study of classrooms (Koivusaari 2002) and committee meetings (Pagliari and Grimshaw 2002)—in fact, in any situation where people interact with one another.

Stewart (1984) audiotaped 140 doctor-patient interactions in the offices of 24 family physicians and assessed the interactions with Bales's Interaction Process Analysis. Ten days later, Stewart interviewed the patients at their homes to assess satisfaction and compliance. That is, were the patients satisfied with the care they'd gotten and were they taking the pills they'd been told to take? Sure enough, when physicians are coded as engaging in many patient-centered behaviors, patients report higher compliance and satisfaction.

One of the best things about the Interaction Process Analysis system is that any act of communication can be identified as being one of those 12 categories in figure 15.2, and the 12 categories are recognizable in many cultures around the world. A more detailed outline for coding interpersonal relations was developed by Bales and Cohen (1979). A complete course on how to use their system is available in their book, aptly titled *SYMLOG*, which stands for “systematic multiple level observation of groups.”

Make no mistake about this: Continuous monitoring is tough to do. It takes several months of intensive training for observers to become adept at using complex coding schemes. In the 1920s, Rudolf von Laban developed a system of 114 signs with which to record dance or any other set of human movements

PROBLEM AREAS		OBSERVATION CATEGORIES
Positive Reactions	A	1 Shows solidarity, raises other's status, gives help, rewards
		2 Shows tension release, jokes, laughs, shows satisfaction
		3 Agrees, shows passive acceptance, understands, concurs, complies
Attempted Answers	B	4 Gives suggestions, direction, implying autonomy for other
		5 Gives opinion, evaluation, analysis, expresses feelings, wishes
		6 Gives orientation, information, repeats, clarifies, confirms
Questions	C	7 Asks for orientation, information, repetition, confirmation
		8 Asks for opinion, evaluation, analysis, expression of feeling
		9 Asks for suggestions, direction, possible ways of action
Negative Reactions	D	10 Disagrees, shows passive rejection, formality, withholds help
		11 Shows tension, asks for help, withdraws out of field
		12 Shows antagonism, deflates other's status, defends or asserts self

Figure 15.2. Categories for direct observation.

SOURCE: Adapted from R. F. Bales, *Social Interaction Systems: Theory and Measurement*, p. 165. © 2001, Transaction Publishers. Reprinted by permission.

(see Lange 1975). If you are trying to understand the meaning of complex human body movement, you've got to start with some way to record the basic data, like any other text. Brenda Farnell used Laban's script (known as **labanotation**) in her meticulous research on Plains Indian sign language (1995) and has written extensively about how the system can be used in the anthropo-

logical study of human movement in general (Farnell 1994, 1996; Farnell and Graham 1998).

These days, behavioral coding in psychology and ethology is being done with handheld computers. This is being facilitated by software that lets you program any key to mean “initiates conversation,” “reciprocates affect,” or whatever. (See Ice [2004] for a detailed review of this technology and appendix F.) As it becomes easier for fieldworkers to observe and code behavior at the same time, I think we’ll see renewed interest in continuous monitoring and in the use of complex coding schemes like labanotation. (One of the problems in the use of direct observation is the need for reliable coding by several researchers of the same data. We’ll take up measures of intercoder reliability in chapter 17 on text analysis. The problem of testing intercoder reliability is the same, whether you’re coding text or behavior.)

Comparative Research: The Six Culture Study

Broad, general coding schemes are particularly useful for comparative research. Whether you’re comparing sessions of psychotherapy groups, interaction sessions in laboratory experiments, or the natural behavior of people in field studies, using a common coding scheme really pays off because you can make direct comparisons across cases and look for generalizations.

The most important comparative study of children ever was run by Beatrice and John Whiting between 1954 and 1956. In the Six Culture Project, field researchers spent from 6 to 14 months in Okinawa, Kenya, Mexico, the Philippines, New England, and India. They made a total of some 3,000 5-minute (continuous monitoring) observations on 67 girls and 67 boys between the ages of 3 and 11.

Observations were limited to just 5 minutes because they were so intense, produced so much data, and required so much concentration and effort that researchers would have become fatigued and lost a lot of data in longer sessions. The investigators wrote out, in clear sentences, everything they saw children doing during the observation periods. They also recorded data about the physical environment and others with whom children were interacting.

The data were sent from the field to Harvard University for coding according to a scheme of 12 behavior categories that had been worked out in research going back some 15 years before the Six Culture Study began. The behavioral categories included: seeks help, seeks attention, seeks dominance, suggests, offers support, offers help, acts socially, touches, reprimands, assaults sociably, assaults not sociably, symbolic aggression (frightens, insults, threatens with gesture, challenges to compete). (Full details on the use of the Whiting

scheme are published in Whiting et al. [1966]. See Whiting and Whiting [1973] for a discussion of their methods for observing and recording behavior.)

On average, every 10th observation was coded by two people, and these pairs of “coding partners” were rotated so that coders could not slip into a comfortable pattern with one another. Coders achieved 87% agreement on children’s actions; that is, given a list of 12 kinds of things a child might be doing, coders agreed 87% of the time. They also agreed 75% of the time on the act that precipitated a child’s actions and 80% of the time on the effects of a child’s actions (Whiting and Whiting 1975:55).

The database from the Six Culture Study consists of approximately 20,000 recorded acts, for 134 children, or about 150 acts per child, on average.

Very strong conclusions can be drawn from this kind of robust database. For example, Whiting and Whiting (1975:179) note that nurturance, responsibility, success, authority, and casual intimacy “are types of behavior that are differentially preferred by different cultures.” They conclude that “these values are apparently transmitted to the child before the age of six.” They found no difference in amount of nurturant behavior among boys and girls 3–5 years of age. After that, however, nurturant behavior by girls increases rapidly with age, while boys’ scores on this trait remain stable.

By contrast, reprimanding behavior starts out low for both boys and girls and increases with age equally for both sexes, across six cultures. The older the children get, the more likely they are to reprimand anyone who deviates from newly learned cultural rules. “Throughout the world,” the Whitings conclude, “two of the dominant personality traits of children between seven and eleven are self-righteousness and bossiness” (1975:184). Anyone who grew up with an older sibling already knows that, but the Whitings’ demonstration of this cross-cultural fact is a major scientific achievement.

Using Video for Continuous Monitoring

Even with a fixed coding scheme, an observer in a CM situation has to decide among alternatives when noting behavior—whether someone is acting aggressively, or just engaging in rough play, for example. Recording behavior on film or video lets several analysts study the behavior stream and decide at leisure how to code it. It also makes your data available for coding by others, now and in the future. (Human ethologists, like Irenäus Eibl-Eibesfeldt [1989], have amassed hundreds of miles of film and videotape of ordinary people doing ordinary things across the world.)

In the 1970s, Marvin Harris and his students installed videotape cameras in

the public rooms of several households in New York City. Families gave their permission, of course, and were guaranteed legal control over the cameras during the study and of the videotapes after the cameras were removed. Teams of observers monitored the equipment from remote locations. Later, the continuous verbal and nonverbal data were coded to study regularities in interpersonal relations in families.

Anna Lou Dehavenon (1978), for example, studied two black and two white families for 3 weeks and coded their nonverbal behavior for such things as compliance with requests and the distribution and consumption of foods in the households. Dehavenon's data showed that the amount of authoritarianism in the four families correlated perfectly with income differences. The lower the family income, the more superordinate behavior in the home (1978:3).

One would hypothesize, from participant observation alone, that this was the case. But *testing* this kind of hypothesis requires the sort of quantified data that straightforward, direct observation provides. (See Sharff [1979] and Reiss [1985] for two more studies of households using the Harris videotapes.)

As video cameras have gotten smaller, easier to use, and less expensive, more field researchers have been using this technology for close examination of behavior streams. Brigitte Jordan, for example, used videotape in her study of birthing events across cultures (1992a; also Jordan and Henderson 1993).

Video is now used routinely in studies by anthropologists of consumer behavior. Observers at Planmetrics, a marketing research firm, videotaped 70 volunteer parents, for over 200 hours, as the volunteers diapered their babies. The research was done on contract with Kimberly-Clark, manufacturer of "Huggies," a brand of disposable diapers. The cameras were not hidden, and after a while people just went about their business as usual, according to Steven Barnett, the anthropologist who led the study.

Close observation showed that many parents could not tell whether their babies needed a diaper change, so the researchers recommended that the diapers contain an exterior chemical strip that changed color when the baby was wet. The observers also noticed that parents were powdering their babies' legs and that parents were treating the red marks left by the diaper gathers as if the marks were diaper rash. The firm recommended that the gathers be redesigned so that there would be no more red marks (Associated Press, 1 October 1985; *The Wall Street Journal*, September 4, 1986:29).

CM and Reactivity

Finally, there are two ways to lower reactivity in continuous monitoring. One of them is participant observation. Once you've built up rapport and trust in a field situation, people are less likely to change their behavior when you're

around. Even if they do change their behavior, you're more likely to notice the change and take that fact into account.

The second way to lower reactivity is training. We can't eliminate observer bias entirely, but lots and lots of evidence shows that training helps make people better—more reliable and more accurate—observers (Kent et al. 1977; Hartmann and Wood 1990). We do the best we can. Just because a “perfectly aseptic environment is impossible,” Clifford Geertz reminds us (paraphrasing the economist Robert Solow), doesn't mean we “might as well conduct surgery in a sewer” (Geertz 1973:30).

Joel Gittelsohn and his coworkers (1997) tested the effects of participant observation and training on reactivity in their study of child-care practices in rural Nepal. Over the course of a year, 10 trained fieldworkers observed behavior in 160 households. Each home was visited seven times. Except for a 3–4 hour break in the middle of the day, the fieldworkers observed a focal child, 2–5 years of age, and all the caregivers of that child, from 6:00 A.M. until 8:00 P.M. This study, then, involved both children and adults.

The observers coded for over 40 activities, including health-related behaviors, feeding activities, and various kinds of social interactions (punishment, affection, and so on). The rate of some behaviors changed a lot over the course of the year. On average, across 1,101 observations, the number of times per day that a caregiver served food to a child without asking the child if he or she wanted it fell by half.

The observers also coded *each time they were interrupted* by one of the people whom they were observing (and what the interruption was about: light conversation, being asked for favors or medicine). This allowed Gittelsohn et al. to track reactivity across the seven household visits. Reactivity was noticeable during the first visit and then fell off dramatically. This study shows clearly that: (1) reactivity exists, and (2) it goes away quickly when indigenous observers stay on the job over time (Gittelsohn et al. 1997).

Spot Sampling and Time Allocation Studies

Instantaneous spot sampling, or **time sampling** was developed in behavioral psychology, in the 1920s, and is widely used in ethology today. In **time allocation** (TA) studies, which are based on time sampling, an observer appears at randomly selected places, and at randomly selected times, and records what people are doing when they are first seen (Gross 1984).

The idea behind the TA method is simple and appealing: If you sample a sufficiently large number of representative acts, you can use the percentage of

times people are seen doing things (working, playing, resting, eating) as a proxy for the percentage of *time* they spend in those activities.

Charles Erasmus used spot sampling in his study of a Mayo Indian community in northern Mexico (1955). As Erasmus and his wife went about the village, investigating “various topics of ethnographic interest,” they took notes of what people were doing at the moment they encountered them. They did not use a representative sampling strategy but they were very systematic in their recording of data.

Individual charts were made for each man, woman, and child in the village, and on those charts were noted the page numbers from the field log where the activity descriptions were to be found. These page numbers were recorded on the charts according to the hours of the day when the observations were made. Thus, the individual charts served as indexes to the field log as well as a means of making sure that equal attention was being given to all families at all hours of the day. Periodic examination of the charts showed which households and which hours of the day were being neglected, so that visits about the community could be planned to compensate for these discrepancies. (Erasmus 1955:325)

It’s difficult to top this research for sheer elegance of design and the power of the data it produced. In the 3 months from July to September 1948, the Erasmuses made about 5,000 observations on 2,500 active adults, 2,000 children, and 500 elders in the community. From those observations, Erasmus demonstrated that men in the village he studied spent about the same time at work each day as did semiskilled workers in Washington, D.C. At the time, Melville Herskovits was trying to combat the racist notion that primitive and peasant peoples are lazy and unwilling to exert themselves. Herskovits’s assertion was vindicated by Erasmus’s TA research.

Reactivity in TA Research

In CM, getting around the reactivity problem involves staying with the program long enough to get people accustomed to your being around. Eventually, people just get plain tired of trying to manage your impression and they act naturally.

In TA research, the trick is to catch a glimpse of people in their natural activities before they see you coming on the scene—before they have a chance to modify their behavior.

Richard Scaglione (1986) did a TA survey of the residents of Upper Neli-gum, a Samakundi Abelam village in the Prince Alexander Mountains of East Sepik Province in Papua New Guinea. “It is not easy,” he says, “for an anthropologist in the field to come upon an Abelam unawares. Since I did not

want to record ‘greeting anthropologist’ as a frequent activity when people were first observed, I often had to reconstruct what they were doing immediately before I arrived” (p. 540).

Monique Borgerhoff-Mulder and Tim Caro (1985) coded the observer’s judgment of whether people saw the observer first, or vice versa, and compared that to whether the Kipsigis (in Kenya) they were studying were observed to be active or idle. People were coded as being idle significantly more often when they spied the observer coming before the observer saw them.

Did people become idle when they saw an observer approaching? Or was it easier for idle people to see an observer before the observer saw them? Borgerhoff-Mulder and Caro found that people who were idle were sitting or lying down much more often than were people who were active. People at rest may be more attentive to their surroundings than those who are working and would be judged more often to have seen the researcher approaching.

Sampling Problems

There are five questions to ask when drawing a sample for a TA study:

1. Who do I watch?
2. Where do I go to watch them?
3. When do I go there?
4. How often do I go there?
5. How long do I spend watching people when I get there? (Gross 1984)

Allen Johnson’s study (1975) of the Machiguenga is instructive. The Machiguenga are horticulturalists in the Peruvian Amazon. They live along streams, in small groups of related families, with each group comprising from about 10 to 30 people, and subsist primarily from slash-and-burn gardens. They supplement their diet with fish, grubs, wild fruits, and occasional monkeys from the surrounding tropical forest. Johnson spent 14 months studying the Machiguenga in the community of Shimaa.

Johnson’s strategy for selecting people to study was simple: Because all travel was on foot, he decided to sample all the households within 45 minutes of his own residence. This produced a convenience sample of 13 households totaling 105 persons. Since the Machiguenga live along streams, each time Johnson went out he walked either upstream or downstream, stopping at a selected household along the route. Which hour of the day to go out, and which houses to stop at were determined by using a table of random numbers, like the one in appendix A of this book.

Thus, Johnson used a nonrandom sample of all Machiguenga households, but he randomized the times that he visited any household in his sample. This sampling strategy sacrificed some external validity, but it was high on internal validity. Johnson could not claim that his sample of households *statistically* represented all Machiguenga households. His 14 months of experience in the field, however, makes his claim for the representativeness of his data credible.

That is, if Johnson's data on time allocation in those 13 households seem to *him* to reflect time allocation in Machiguenga households generally, then they probably do. But we can't be sure. Fortunately, randomizing his visits to the 13 households, and making a lot of observations (3,945 of them, over 134 different days during the 14-month fieldwork period), gives Johnson's results a lot of *internal* validity. So, even if you're skeptical of the external validity of Johnson's study, you could repeat it (in Shima or in some other Machiguenga community) and see whether you got the same results.

Regina Smith Oboler (1985) did a TA study among the Nandi of Kenya. She was interested in differences in the activities of adult men and women. The Nandi, Oboler said, "conceptualize the division of labor as sex segregated. Is this true in practice as well? Do men and women spend their time in substantially different or similar types of activities?" (p. 203).

Oboler selected 11 households, comprising 117 people, for her TA study. Her sample was not random. "Selecting a random sample," she said, "even for one *kokwet* (neighborhood) would have made observations impossibly difficult in terms of travel time" (ibid.:204). Instead, Oboler chose a sample of households that were matched to social and demographic characteristics of the total population and within half an hour walking distance from the compound where she lived.

Oboler divided the daylight hours of the week into 175 equal time periods and gave each period (about two hours) a unique three-digit number. Then, using a table of random numbers, she chose time periods to visit each household. She visited each household four times a week (on different days of the week) during 2 weeks each month and made nearly 1,500 observations on those households during her 9 months in the field.

Oboler found that, for her sample of observations, adult men spend around 38% of their time "in activities that might reasonably be considered 'work' by most commonly used definitions of that term" (ibid.:205). Women in her sample spent over 60% of their time working.

Table 15.2 shows the number of spot observations necessary to estimate the frequency of an activity to within a fractional accuracy. It also tells you how many observations you need if you want to see an activity at least once with 95% probability.

Here's how to read the table. Suppose people spend about 5% of their time

TABLE 15.2
Number of Observations Needed to Estimate the Frequency of an Activity to
within a Fractional Accuracy

True Frequency of Activity	Number of Observations Needed to See the Activity at a Particular Fraction of Accuracy								To See Activities at Least Once with 95% Probability
	0.05	0.10	0.15	0.20	0.30	0.40	0.50		
<i>f</i>									
0.01	152127	38032	16903	9508	4226	2377	1521	299	
0.02	75295	18824	8366	4706	2092	1176	753	149	
0.03	49685	12421	5521	3105	1380	776	497	99	
0.04	36879	9220	4098	2305	1024	576	369	74	
0.05	29196	7299	3244	1825	811	456	292	59	
0.06	24074	6019	2675	1505	669	376	241	49	
0.07	20415	5104	2268	1276	567	319	204	42	
0.08	17671	4418	1963	1104	491	276	177	36	
0.09	15537	3884	1726	971	432	243	155	32	
0.10	13830	3457	1537	864	384	216	138	29	
0.15	8708	2177	968	544	242	136	87	19	
0.20	6147	1537	683	384	171	96	61	14	
0.25	4610	1152	512	288	128	72	46	11	
0.30	3585	896	398	224	100	56	36	9	
0.40	2305	576	256	144	64	36	23	6	
0.50	1537	384	171	96	43	24	15	5	

SOURCE: H. R. Bernard and P. D. Killworth, "Sampling in Time Allocation Research," *Ethnology*, Vol. 32, pp. 211. Copyright © 1993. Reprinted with permission.

eating. This is shown in the first column as a frequency, *f*, of 0.05. If you want to estimate the frequency of the activity to within 20%, look across to the column in the center part of table 15.2 under 0.20. If you have 1,825 observations, and your data say that people eat 5% of the time, then you can safely say that the true percentage of time spent eating is between 4% and 6%. (Twenty percent of 5% is 1%; 5%, plus or minus 1%, is 4%–6%. For the formula used to derive the numbers in table 15.2, see Bernard and Killworth 1993.)

Suppose you do a study of the daily activities of families in a community and your data show that men eat 4% of the time and women eat 6% of the time. If you have 300 observations, then the error bounds of the two estimates overlap considerably (about 0.02–0.06 for the men and 0.04–0.08 for the women).

You need about 1,800 observations to tell whether 0.06 is really bigger than

0.04 comparing across groups. It's the same for activities: If women are seen at leisure 20% of their time and caring for children 25% of their time, then you need 1,066 observations to tell if women really spend more time caring for children than they do at leisure.

Oboler had 1,500 observations. It is clear from table 15.2 that her findings about men's and women's leisure and work time are not accidents. An activity seen in a sample of just 256 observations to occur 40% of the time can be estimated actually to occur between 40%, plus or minus 15% of 40%, or between 34% and 46%. Since men are seen working 38% of the time and about half of Oboler's 1,500 observations were of men, her finding is solid.

Nighttime Sampling

Most time allocation studies are done during the daylight hours, between 6 A.M. and 7–8 P.M. In Johnson's case, this was explicitly because "travel after dark is hazardous, and because visiting at night is not encouraged by the Machiguenga" (Johnson 1975:303).

Of course, we know that life doesn't stop when the sun goes down. Kenneth Good spent 13 years with the Yanomami in the Venezuelan Amazon. The Yanomami, Good reports, sit around at night and plan the next day's hunt—not in whispers, but in full volume. When the mood strikes, a Yanomami man might sit in his hammock and give a speech, to everyone within earshot. "A Yanomami night," says Good, "was like another day. . . . Like the Yanomami, I'd spend eleven hours in my hammock at night to get seven or eight hours of actual sleep" (Good 1991:41–42).

Richard Scaglione (1986) did nighttime spot observations in his study of the Abelam in New Guinea. In 1983, when Scaglione did his study, there were 350 people in the village, living in 100 households. Scaglione randomly selected 2 households each day, and visited them at randomly selected times, throughout the day *and night*.

Scaglione didn't get much sleep during the month that he did this work, but his findings were worth the sacrifice. He coded his 153 observations into 13 categories of activities: sleeping, gardening, idle, cooking and food preparation, ritual, visiting, eating, hunting, construction, personal hygiene, child care, cleansing and washing, and craftwork. Only 74% of Abelam activities *during nighttime hours* were coded as "sleeping." Seven of the nine observations that he coded as "ritual" occurred after dark. Half of all observations coded as "hunting" occurred at night, and six out of eight observations coded as "visiting" were nocturnal.

Had he done his TA study only during the day, Scaglione would have overestimated the amount of time that Abelam people spend gardening by about a

fourth. His data show that gardening takes up about 26% of the Abela's daylight hours, but only 20% of their total waking time in each 24-hour period.

Of course, it may not always be possible to conduct TA studies at night. Johnson, you'll remember, made a point of the fact that the Machiguenga discourage nighttime visiting. Scaglione, on the other hand, worked among a people who "go visiting at unusual hours, even when their prospective host is likely to be sleeping."

Scaglione, in fact, rather enjoyed showing up at people's houses during odd hours in 1983. He had done his doctoral research in the same village in 1974–1975. In those days, he says, "I was still quite a novelty. . . . I was frequently awakened by hearing '*Minoa, mine kwak?*' ('Hey, you, are you sleeping?'). This study allowed me to return old favors by visiting people in the late night hours to be sure *they* were sleeping" (ibid.:539). (For more examples of TA studies, see Bock 2002 and Umezaki et al. 2002.)

Coding and Recording Time Allocation Data

Sampling is one of two problems in TA research. The other is measurement. How do we know that when Oboler recorded that someone was "working," we would have recorded the same thing? If you were with Johnson when he recorded that someone was engaged in "hygiene behavior," would you have agreed with his assessment? Every time? You see the problem.

It gets even more thorny. Suppose you work out a coding scheme that everyone agrees with. And suppose you train other observers to see just what you see. (Rogoff [1978] achieved a phenomenal 98% interobserver agreement in her study of 9 year olds in Guatemala.) Or, if you are doing the research all by yourself, suppose you are absolutely consistent in recording behaviors (i.e., you never code someone lying in a hammock as sleeping when they're just lounging around awake).

Even if all these reliability problems are taken care of, what about observation validity? What do you do, for example, when you see people engaged in multiple behaviors? A woman might be holding a baby and stirring a pot at the same time. Do you code her as engaged in child care or in cooking? (Gross 1984:542). If someone saw that you were lying down reading, and you were studying for an exam, should they record that you were working or relaxing?

Do you record all behaviors? Do you mark one behavior as primary? This last question has important implications for data analysis. There are only so many minutes in a day, and the percentage of people's time that they allocate to activities has to add up to 100%. If you code multiple activities as equally important, then there will be more than 100% of the day accounted for. Most TA researchers use their intuition, based on participant observation, to decide

which of the multiple simultaneous activities they witness to record as the primary one and which as secondary.

The best solution is to record *all* possible behaviors you observe in the order of their primacy, according to your best judgment at the time of observation. Use a check sheet to record behaviors. Use a separate check sheet for each observation you make. This can mean printing up 1,000 sheets for a TA study, and hauling them home later.

This is one of two good reasons to use handheld computers to record observational data. Programs are now available for these devices that give you the equivalent of an electronic check sheet and let you watch behavior and code it on the spot (see appendix F).

The other good reason to give up paper check sheets if you can is that, eventually, you have to type in the data anyway, transferring them from paper to computer so you can run statistical analyses. Each time you record behavioral data—as you watch it and as you transfer it from paper to computer—there is a chance of error. That’s just how it is. So eliminating one step in the data management process cuts down on errors in data.

If you must hand-code your original observations, then enter the data into a laptop while you’re still in the field, as a precaution against loss of the original data sheets. Be paranoid about data. Those horror stories you’ve heard about lost data? They’re true. (Read M. N. Srinivas’s account [1979:xiii] of how he lost all three copies of his field notes, compiled over a period of 18 years, in a fire at Stanford.)

Experience Sampling (ES)

In **experience sampling** (ES), people carry around little beepers. When the beeper goes off, they fill out a questionnaire about what they’re doing, or who they’re with, or what they’re feeling at the moment. Some researchers ask informants to jot an entry into a diary or talk about their actions, feelings, and surroundings into a small digital or tape recorder, and some researchers even ask informants to call a number and report, live, to a telephone interviewer (Kubey et al. 1996).

ES offers two big advantages. First, it combines the power of random spot checks with the relative ease of having people report on their own behavior. Csikszentmihalyi and Larson (1987) demonstrated the reliability of ES in a number of studies. Validity is another matter, of course, but when people record or talk about what they’re doing and how they’re feeling on the spot, this should lessen the inherent inaccuracy of recall data. The other advantage is that, working on your own, you can only be in one place at a time, but with

those beepers, you can collect spot-observation data from lots of people at once.

In anthropology, Garry Chick used ES in his study (1994) of a small machine-tool company in Pennsylvania. Chick wanted to test Marx's theory that automation would eliminate the need for skilled labor by turning complex tasks into a series of routine steps that could be performed by drones. Drones should find their work boring, unsatisfying, and unimportant.

There were 11 machinists, all men, in the small shop. Four of them worked on traditional, manually controlled lathes in turning out machine tools. Three worked only on a newly installed, computer-controlled lathe (you program it to do a job and then you more-or-less stand by while it executes your instructions). And four worked on both types of lathes, depending on the job.

Each man was beeped 30 times—six times a day over a 5-day week—and at each beep, he stopped what he was doing within 10 minutes and filled out a 2-minute questionnaire. (Each man kept a spiral-bound booklet of questionnaires handy in the shop.) There should have been 330 questionnaires (11 men filling out 30 each), but one man was beeped while he was in the bathroom and one was beeped while he was filling out a questionnaire (having been beeped two minutes before). That's what happens when you collect data at random points, but random means random and no tinkering with it.

For each of the experience samples, workers indicated what they were doing. If they were operating a machine (or more than one), they indicated which one(s). Then they answered 14 attitude questions. Here they are:

On a scale from 1–10, where 1 means “definitely not” and 10 means “definitely yes,” what I was doing:

1. was enjoyable.
2. was interesting.
3. was complex/technical.
4. was fun.
5. was under my control.
6. was monotonous.
7. was machine paced.
8. was tricky.
9. held my attention.

On a scale from 1–10, where 1 means “definitely not” and 10 means “definitely yes,” at the time I was signaled, I was:

10. pressed for time.
11. working on my own.
12. thinking about things other than work.
13. doing something that I felt was important.
14. doing something that required a lot of skill.

The results were mixed. The men found working on the computer-controlled machines more interesting, more satisfying, and more important than working on the manual machines, and they found programming the computer-controlled machines even more interesting than running them. But these machinists also felt more in control when they worked on manual lathes.

If you use self-administered questionnaires, of course, you need literate informants to do experience sampling. But, as Chick says, “with more and more anthropological research in modern and modernizing societies, experience sampling can be a valuable addition to the anthropologist’s tool kit” (1994:6). And if you give people little tape recorders or digital recorders, you may be able to use the ES method with nonliterate populations. (For more on the experience sampling method, see Scollon et al. [2005]; and see Kahneman et al. [2004] for a method that duplicates experience sampling through direct interviews rather than with beeper interrupts.)

Combining Continuous Monitoring and Spot Sampling

The difference between CM and spot sampling is analogous to the difference between ethnography and survey research. With ethnography, you get information about process; with survey research, you get data that let you estimate parameters for a population. Spot sampling is used in TA research precisely because the goal is to estimate parameters—like how much time, on average, women spend cooking, or men spend throwing pots, or children spend laid up ill at home. If you want to know the ingredients of *mafongo* (a dish native to Puerto Rico) and the order in which they are added, you have to watch continuously as someone makes it. (Making it yourself, as part of participant observation, produces embodied knowledge, yet a third kind of information.)

Robin O’Brian (1998) combined CM and spot sampling in her study of Mayan craftswomen in Chiapas, Mexico. The women sold their crafts to tourists at a local market. Several times a week, O’Brian went through the market (entering from a randomly selected spot each time) and coded what every woman craft seller was doing, using a check sheet adapted from Johnson (1987) standardized time allocation activity codes.

O’Brian also did continuous monitoring for 3 hours of 15 women, and these two kinds of data produced more information than either kind alone. Her aggregate, spot-sampling data showed that the women spent 82% of their time waiting for tourists to buy something. The women weren’t just sitting around, though. They spent 17% of their waiting time producing more crafts (doing macramé or embroidery or hand-spinning wool) and another 17% eating,

cleaning children, or nursing. The CM data showed that women were interrupted in their productive activities 36% of their waiting time (that is, 36% of 82%, or 30% of their entire day) and that the interruptions were as likely to be responding to their children's needs as they were to be selling to tourists.

Pluses and Minuses of Direct Observation

On balance, direct observation provides much more accurate results about behavior than do reports of behavior. Ricci et al. (1995) studied 40 people in Kalama, a peri-urban village about 15 miles north of Cairo. One day, a set of trained observers watched the 40 participants for 2.5 hours. The observers noted whether people were engaged in any of 17 activities and how long each person spent at each activity.

The next day, the participants were asked to recall, sequentially, what they had done the entire day before, and how long they had spent at each activity. The interviewers did not mention any of the 17 activities, but they tried to improve respondent recall by asking about activities before and after locally significant time markers, like call to prayer. Ten of the 40 were toddlers, so Ricci et al. focused on the recall data of the 24 adults and the six school-age children.

Ricci et al. were very forgiving in their analysis. Informants were scored as being correct if they could recall an activity at all and say correctly whether it had been done in the morning or afternoon observation period (9:00–11:30 A.M. or 12:30–3:00 P.M.). Ricci et al. scored only errors of omission (leaving out activities) and threw out errors of commission (inventing activities that had not been observed).

And informants *still* got it wrong—a lot. Across men and women, across agricultural and nonagricultural households, informants got it wrong, on average, 56% of the time. Five of the six women who had been observed breast-feeding failed to report that activity the next day; 13 of the 15 women who had been observed washing clothes failed to report that activity the next day.

If you want to know whether, say, caring for animals happens more often than, say, gathering fuel, then self-reports might be enough. But if you want to know *how often* those behaviors actually occur, then nothing short of direct observation will do.

I don't want to give the impression, however, that direct observation data are automatically accurate. Lots of things can clobber the accuracy of directly observed behavior. Observers may be biased by their own expectations of what they are looking for or by expectations about the behavior of women or men or any ethnic group (Kent et al. 1977; Repp et al. 1988).

You may feel awkward about walking around with a clipboard (and perhaps a stopwatch) and writing down what people are doing—or with beeping people and asking them to interrupt what they're doing to help you get some data. This is a reasonable concern, and direct observation is not for everyone. It's not a detached method, like sending out questionnaires and waiting for data to be delivered to your doorstep.

It is not a fun method, either. Hanging out, participating in normal daily activities with people, and writing up field notes at night is more enjoyable than monitoring and recording what people are doing.

But many fieldworkers find that direct observation allows them to address issues that are not easily studied by any other method. Grace Marquis (1990) studied a shantytown in Lima, Peru. Children in households that kept chickens were at higher risk for getting diarrhea than were other children. The chickens left feces in the homes, and the feces contained an organism that causes diarrhea. Continuous monitoring showed that children touched the chicken droppings and, inevitably, touched their mouths with their hands. It was hard, tedious work, but the payoff was serious.

Direct observation is time consuming, but random spot-checking of behavior is a very cost effective and productive way to use *some* of your time in any field project. When you're studying a group that has clear boundaries (a village, a hospital, a school), you can get very fine-grained data about people's behavior from a TA study, based on random spot checks. More importantly, as you can see from table 15.2, with proper sampling you can generalize to large populations (whole school districts, an entire aircraft manufacturing plant, even cities) from spot checks of behavior, in ways that no other method allows.

You may be concerned that a strictly observational approach to gathering data about human behavior fails to capture the *meaning* of data for the actors. This, too, is a legitimate concern. A classic example is Geertz's (1973:6–7) observation that a wink can be the result of getting a speck of dust in your eye or a conscious act of conspiracy. And that's just a wink. People can engage in any of thousands of behaviors (skipping a class, wearing a tie, having their navel pierced . . .) for many, many different reasons. Knowing the meaning of behavior to others is essential to understanding it ourselves.

On the other hand, one of our most important goals in science is to constantly challenge our own ideas about what things mean. That's how theories develop, are knocked down, and gain in their power to explain things. Why shouldn't we also challenge the theories—the explanations—that the people we study give us for their own behavior?

Ask people who are coming out of a church, for example, why they just spent 2 hours there. Some common responses include “to worship God,” “to

be a better person,” “to teach our children good values.” Hardly anyone says “to dress up and look good in front of other people,” “to meet potential golf partners for later this Sunday afternoon,” “to maximize my ability to meet potential mates whose ethnic and social backgrounds are compatible with my own.” Yet, we know that these last three reasons are what *some* people would say if they thought others wouldn’t disapprove.

Finally, you may have some qualms about the ethics of obtrusive observation. It cannot be said too often that *every single data collection act* in the field has an ethical component, and a fieldworker is obliged every single time to think through the ethical implications of data collection acts. Personally, I have less difficulty with the potential ethical problems of obtrusive, reactive observation than I do with any other data collection method, including participant observation. In obtrusive observation, people actually *see* you (or a camera) taking down their behavior, and they can ask you to stop. Nothing is hidden.

In participant observation, we try to put people at ease, make them forget we’re really listening hard to what they’re telling us, and get them to “open up.” We ask people to take us into their confidence and we are handed the responsibility for not abusing that confidence.

But the method that presents the *most* ethical problems is unobtrusive, *non-reactive* direct observation.

Unobtrusive Observation

Disguised field observation is the ultimate in participant observation—you join, or pretend to join, some group and secretly record data about people in the group.

In 1960, John H. Griffin, a white journalist, went through some drug treatment to temporarily turn his skin black. He traveled the southern United States for about a month, taking notes on how he was treated. His book, *Black Like Me* (1961) was a real shocker. It galvanized a lot of support by Whites in the North for the then fledgling Civil Rights movement. Clearly, Griffin engaged in premeditated deception in gathering the data for his book. But Griffin was a journalist; scientists don’t deceive their informants, right?

Pseudopatients and Simulated Clients

Wrong. Samuel Sarkodie, an M.A. student in medical sociology at the University of Legon, in Ghana, spent 3 days in a rural hospital in 1994 as a pseudopatient with a false case of malaria. The hospital staff were in on the

study—they had been recruited by Sarkodie’s supervisor, Sjaak van der Geest, a Dutch anthropologist who works in Ghana—and Sarkodie wrote a detailed report for the hospital. Presumably, the report was helpful, but Sarkodie’s fellow patients were duped (van der Geest and Sarkodie 1998).

Twenty years earlier, David Rosenhan recruited seven confederates who, like him, checked themselves into mental hospitals and took surreptitious notes about how they were treated. They gave false names and occupations (they couldn’t very well mention their real occupations since three of them were psychologists and one was a psychiatrist), and reported hearing voices. One was diagnosed as manic-depressive, and the rest as schizophrenics, and all were admitted for treatment.

This was tough work. The pseudopatients were not allowed to divulge what they were up to just because they were tired of (or exasperated with) the experiment. The only way out was to be diagnosed by the hospital staff as ready for release. It took between 1 week and 7 weeks of confinement to achieve this, and when they were released, the pseudopatients were all diagnosed with “schizophrenia in remission” or as “asymptomatic” or as “improved” (Rosenhan 1973, 1975).

Rosenhan’s field experiment made clear the power of labeling: Once you are diagnosed as insane, people treat you as insane. Period. Some of the genuine inmates at the hospitals saw through the charade, but none of the staff ever did.

Are these pseudopatient studies ethical? The simulated client method has been used in dozens of studies to evaluate the performance of physicians, pharmacists, family-planning clinics, and other health care providers in developing nations. (See Madden et al. [1997] for a review of these studies; and see Tuladhar et al. [1998], Katz and Naré [2002], Chalker et al. [2004], and Marsh et al. [2004] for recent examples.) And fake clients—men and women, black, white, and Hispanic—are sent out by U.S. government agencies regularly to apply for jobs, to rent apartments, or to buy homes and to uncover discrimination (Sharpe 1998). The U.S. Supreme Court has ruled that this practice is legal in the pursuit of fair housing (Ayres 1991:823), and the Equal Employment Opportunity Commission uses data from these field experiments to sue offending businesses.

People across the political spectrum have quite different ideas about whether this is just a dose of the same medicine that offenders dish out (which seems fair), or entrapment (which seems foul). Does this mean that ethics are simply a matter of political orientation and opinion? In the abstract, most people answer this question with a strong “no.” When things get concrete—when the fortunes and reputations of real people are at stake—the answer becomes less clear for many people.

But if you think deceiving landlords or realtors or the staff of mental hospitals is something, read on.

The Tearoom Trade Study

Without telling people that he was studying them, Laud Humphreys (1975) observed hundreds of homosexual acts among men in St. Louis, Missouri. Humphreys's study produced very important results. The men involved in this tearoom trade, as it is called, came from all walks of life, and many were married and living otherwise straight lives. Humphreys made it clear that he did not engage in homosexual acts himself, but played the role of the "watch queen," or lookout, warning his informants when someone approached the restroom. This deception and unobtrusive observation, however, did not cause the storm of criticism that accompanied the first publication of Humphreys's work in 1970.

That was caused by Humphreys having taken his research a step further. He jotted down the license plate numbers of the men who used the restroom for quick, impersonal sex, and got their names and addresses from motor vehicle records. He waited a year after doing his observational work, and then, on the pretext that they had been randomly selected for inclusion in a general health survey, he interviewed 100 of his research subjects in their homes.

Humphreys was careful to change his car, his hairstyle, and his dress. According to him, his informants did not recognize him as the man who had once played watch queen for them in public toilets. *This* is what made Humphreys' research the focus of another debate, which is still going on, about the ethics of nonreactive field observation.

Five years after the initial study was published, Humphreys himself said that he had made a mistake. He had endangered the social, emotional, and economic lives of people he studied. Had his files been subpoenaed, he could not have claimed immunity. He decided at the time that he would go to jail rather than hurt his informants (Humphreys 1975).

Humphreys was an ordained Episcopal priest who had held a parish for more than a decade before going to graduate school. He was active in the Civil Rights movement in the early 1960s and spent time in jail for committing crimes of conscience. His credentials as an ethical person, conscious of his responsibilities to others, were in good order. Everyone associated with him agreed that Humphreys was totally committed to protecting his informants.

But listen to what Arlene Kaplan Daniels had to say about all this, in a letter to Myron Glazer, a sociologist and ethnographer:

In my opinion, no one in the society deserves to be trusted with hot, incriminating data. Let me repeat, *no one*. . . . We should not have to rely on the individual

strength of conscience which may be required. Psychiatrists, for example, are notorious gossipers [about their patients]. . . . O. K., so they mainly just tell one another. But they *sometimes* tell wives, people at parties, you and me. [Daniels had done participant observation research on psychiatrists.] And few of them would hold up under systematic pressure from government or whatever to get them to tell. . . . The issue is not that a few brave souls *do* resist. The issue is rather what to do about the few who will not. . . . There is *nothing* in our training—any more than in the training of psychiatrists, no matter what they say—to prepare us to take up these burdens. (quoted in Glazer 1975:219–20; emphasis in original)

Researchers who conduct the kinds of studies that Humphreys did, invoke several arguments to justify the use of deception.

1. It is impossible to study such things as homosexual encounters in public restrooms in any other way.
2. Disguised field observation is a technique that is available only to researchers who are physically and linguistically indistinguishable from the people they are studying. To use this technique, you must be a member of the larger culture. There is, therefore, no real ethical question involved, other than whether you, as an individual, feel comfortable doing this kind of research.
3. Public places, like restrooms, are, simply, public. The counterargument is that people have a right to expect that their behavior in public toilets will not be recorded, period. (Koocher 1977)

Sechrest and Phillips (1979) take a middle ground. They say that “public behavior should be observable by any means that protect what might be called ‘assumed’ privacy, the privacy that one might expect from being at a distance from others or of being screened from usual views” (p. 14). Casual observation is fine, but the use of telescopes, listening devices, or peepholes would be unethical.

My own position is that the decision to use deception is up to you, provided that the *risks of detection are your own risks and no one else’s*. When Jack Weatherford (1986) took a job as manager of a porn shop in Washington, D.C., the people who came to the store to watch the movies or connect with prostitutes didn’t know they were being studied by a participant observer, but neither were they in any danger that their identities would be divulged. And similarly, when Wendy Chapkis became a licensed massage therapist and became a participant observer in her secret research on prostitution (1997), she assumed risks, but the risks were hers. If detection risks harm to others, then don’t even consider disguised participant observation. Recognize, too, that it may not be possible to foresee the potential harm that you might do using disguised observation. This is what leads scholars like Kai Erikson (1967,

1996) to the conclusion that research that requires deception is never justified. (For more on deception in participant observation, see Rynkiewick and Spradley 1976, Cassell 1982, Bulmer 1991, Brymer 1998, and Lauder 2003.)

Grades of Deception

But is all deception equally deceitful? Aren't there **grades of deception**? In the 1960s, Edward Hall and others (Hall 1963, 1966; Watson and Graves 1966) showed how people in different cultures use different "body language" to communicate—that is, they stand at different angles to one another, or at different distances when engaging in serious vs. casual conversation. Hall called this different use of space **proxemics**. He noted that people learn this proxemic behavior as part of their early cultural learning and he hypothesized that subcultural variations in spatial orientation often leads to breakdowns in communication, isolation of minorities, and so on.

This seminal observation by an anthropologist set off a flurry of research that continues to this day (for a review, see Farnell 1999). Early on, Aiello and Jones (1971) studied the proxemic behavior of middle-class white and lower-class Puerto Rican and black schoolchildren. They trained a group of elementary schoolteachers to observe and code the distance and orientation of pairs of children to one another during recess periods.

Sure enough, there were clear cultural and gender differences. White children stand much farther apart in ordinary interaction than do either black or Puerto Rican children. The point here is that the teachers were natural participants in the system. The researchers trained these natural participants to be observers in order to cut out any reactivity that outsiders might have caused in doing the observation.

For his Master's degree, Mark House studied shoppers in Nizhny Novgorod, a city on the Volga River about 400 miles east of Moscow in the Volga region. As he followed shoppers around, House held his little tape recorder to his ear, as if he were talking on a cell phone, and dictated notes. Scherer (1974) studied pairs of children in a schoolyard in Toronto. He used only lower-class black and lower-class white children in his study, to control for socioeconomic effects. Scherer adapted techniques from photogrammetry (making surveys by using photographs). He mounted a camera in a park adjacent to the schoolyard. Using a telephoto lens, he took unobtrusive shots of pairs of children who were at least 30 meters away.

This got rid of the reactivity problem. Then Scherer devised a clever way to measure the average distance between two children and did his analysis on the quantitative data. Scherer found no significant differences in the distance

between pairs of white or black children. (For more on proxemics, see Ardener 1981, Kendon 1981, and Low and Lawrence-Zúñiga 2003.)

Levine (1997) used casual, unobtrusive observation to study the walking speed of people in different size cities and Rotton et al. (1990) tested whether people walk faster in a climate-controlled mall or in open-air shopping. Contrary to popular wisdom, heat didn't slow down urban shoppers of either sex. And Sykes et al. (1990) sat unobtrusively in bars, counting the number of drinks people consumed. Confirming popular wisdom, people drink faster and spend less time in bars when they are in groups of two or more than when they're alone.

I don't consider these field studies of shoppers, children, pedestrians, and drinkers in bars to be unethical. The people being studied were observed in the course of their ordinary activities, out in the open, in truly public places. Despite making unobtrusive observations, or taking surreptitious pictures, the deception involved was passive—it didn't involve “taking in” the subjects of the research, making them believe one thing to get them to do another. I don't think that any real invasion of privacy occurred.

The Micturition Study

You can't say that about the work of Middlemist et al. (1976). They wanted to measure the length of time it takes for men to begin urinating, how long men continue to urinate, and whether these things are affected by how close men stand to each other in public toilets. (*Why* they wanted to know these things is another story.)

At first, the researchers pretended to be combing their hair at the sink in a public toilet at a university. They tracked the time between the sound of a fly being unzipped and urine hitting the water in the urinal as the time for onset; then they noted how long it took for the sound of urine to stop hitting the water in the urinal, and counted this as the duration of each event. They noted whether subjects were standing alone, next to someone, or one or two urinals away from someone.

In general, the closer a man stood to another man, the longer it took him to begin urinating and the shorter the duration of the event. This confirmed laboratory research showing that social stress inhibits relaxation of the urethral sphincter in men, thus inhibiting flow of urine.

Middlemist et al. decided to control the independent variable—how far away another man was from each subject. They placed “BEING CLEANED” signs on some urinals, and forced unsuspecting men to use a particular urinal in a public toilet. Then a confederate stood next to the subject, or one urinal away, or did not appear at all. The observer hid in a toilet stall next to the

urinals and made the measurements. The problem was, the observer couldn't hear flies unzipping and urine hitting the water from inside the stall—so the researchers used a periscopic prism, trained on the area of interest, to make the observations directly.

Personally, I doubt that many people would have objected to the study if Middlemist and his colleagues had just lurked in the restroom and done simple, unobtrusive observation. But when they contrived to make men urinate in a specific place, when they contrived to manipulate the dependent variable (urination time), and, above all, when they got that periscope into the act, that changed matters. This is a clear case of invasion of privacy by researchers, in my view.

In a severe critique of the research, Koocher (1977:120) said that “at the very least, the design seems laughable and trivial.” Middlemist et al. (1977:123) defended themselves, saying that “we believe . . . that the pilot observation and the experiment together constitute an example of well-controlled field research, adequate to test the null hypothesis that closeness has no effect” on the duration of urination among males in public restrooms. Actually, Middlemist et al.'s study *design* was anything but trivial. In fact, it was quite elegant. Whether it was ethical is another matter, as is whether the knowledge they were after was worth doing the experiment at all.

Passive Deception

Passive deception involves no experimental manipulation of informants to get them to act in certain ways. Humphreys's first, strictly observational study (not the one where he used a pretext to interview people in their homes) involved passive deception. He made his observations in public places where he had every right to be in the first place. He took no names down, and there were no data that could be traced to any particular individual. Humphreys observed felonies, and that makes the case more complex. But in my opinion, he had the right to observe others in public places, irrespective of whether those observed believed that they would or would not be observed. What he did with his observations—following people up by tracking them through their license plates—is another matter.

Anthropologists use passive deception all the time. I have spent hours pretending to be a shopper in department stores and have observed mothers who are disciplining their children. I have played the role of a strolling tourist on Mexican beaches (an easy role to play, since that was exactly what I was) and recorded how American and Mexican families occupied beach space. I have surreptitiously clocked the time it takes for people who were walking along the streets of Athens (Greece), New York City, Gainesville (Florida), and

Ixmiquilpan (Mexico) to cover 10 meters of sidewalk at various times of the day. I have stood in crowded outdoor bazaars in Mexico, watching and recording differences between Indians and non-Indians in the amount and kinds of produce purchased.

I have never felt the slightest ethical qualm about having made these observations. In my opinion, passive deception is ethically aseptic. Ultimately, however, the responsibility for the choice of method, and for the practical, human consequences of using a particular method, rests with you, the individual researcher. You can't foist off that responsibility on "the profession," or on some "code of ethics."

Are you disturbed by the fact that Humphreys did his research at all, or only by the fact that he came close to compromising his informants? As you answer that question for yourself, you'll have a better idea of where *you* stand on the issue of disguised field observation.

Behavior Trace Studies

Think of trace studies as behavioral archeology. Do people in different cultures really have a different sense of time? Levine and Bartlett (1984) went to 12 cities in six countries and noted the time on 15 randomly chosen bank clocks in each city. Then they measured the difference between the time shown on the clocks and the time reported by the local telephone company in each city. The most accurate public clocks were Japanese—off by an average of just 34 seconds. U.S. clocks were next (off by an average of 54 seconds), followed by the clocks in Taiwan, England, and Italy (71 seconds, 72 seconds, and 90 seconds, respectively). Indonesia came in last, at 189 seconds.

Here you have hard, archeological evidence of clock-setting behavior across six countries. Real people had set those 15 clocks in each city, and real people were responsible for making sure that the clocks were adjusted from time to time. Levine and Bartlett looked at whether differences in the average deviation of the clocks from the real time predicted differences in the rate of heart disease.

They don't. The country with the lowest rate of heart disease, Japan, has the most accurate clocks and the fastest overall pace of life (as measured by several other indicators). Apparently, according to Levine and Bartlett, it's possible in some cultures to be hard *working* without being hard *driving*.

Sechrest and Flores (1969) recorded and analyzed bathroom graffiti in a sample of men's public toilets in Manilla and Chicago. They wanted to examine attitudes toward sexuality in the two cultures. The results were striking. There was no difference in the percentage of graffiti in the two cities that dealt

with heterosexual themes. But fully 42% of the Chicago graffiti dealt with homosexuality, whereas only 2% of the Manilla graffiti did, showing a clear difference in the two cultures regarding level of concern with homosexuality.

Gould and Potter (1984) did a survey of used-up (not smashed-up) automobiles in five Providence, Rhode Island, junkyards. They calculated that the average use-life of American-made cars is 10.56 years, irrespective of how many times cars change hands. This is a good deal longer than most Americans would guess. Gould also compared use-life against initial cost and found that paying more for a car doesn't affect how long it will last. Interesting and useful findings.

In their classic book on *Unobtrusive Measures*, Webb et al. (1966) identified a class of measures based on erosion. Administrators of Chicago's Museum of Science and Industry had found that the vinyl tiles around an exhibit showing live, hatching chicks needed to be replaced about every six weeks. The tiles around other exhibits lasted for years without having to be replaced. Webb et al. (p. 37) suggested that this erosion measure (the rate of wear on vinyl tiles) might be a proxy for a direct measure of the popularity of exhibits. The faster the tiles wear out, the more popular the exhibit.

Weighing the Evidence

Dean Archer and Lynn Erlich (1985) report a method for studying confidential records when you want to know only aggregate outcomes and don't need data about individuals. They had a hypothesis that sensational crimes (with a lot of press coverage) result in increased sales of handguns. The police would not allow them to see the handgun applications, so they asked a member of the police staff to put the permits into envelopes, by month, for 3 months before and 3 months after a particular sensational crime. Then they weighed the envelopes and converted the weight to handgun applications. To do this, they got a chunk of blank applications and found out how many applications there were per ounce.

The technique is very reliable. The correlation between the estimates of researchers and the actual weights of the envelopes was .99, and in a controlled experiment, researchers were able to tell the difference of just one sheet of paper in 15 out of 18 tries. Real data can be messy, though. Lots of handgun applications have addenda attached, for example. The correlation between researchers' estimates and the true number of handgun applications across 6 months was .94.

Archer and Erlich suggest that the weight method can be used to study things like drunk driving arrests, the influx of psychiatric patients to a clinic,

the number of grievance filings in a company, the number of abortion referrals, and the number of complaints against agencies.

The Garbage Project

The largest and most important trace measure research ever attempted is the ongoing Garbage Project, headed by archeologist William Rathje at the University of Arizona. Since 1973, Rathje and his associates have studied the current consumer behavior patterns in Tucson, Arizona, by analyzing the garbage from a representative sample of residents.

In 1988, about 6,000 residents of Tucson were sent flyers, explaining that they were selected to be part of a study of recycling behavior. Their garbage would be studied, the flyer explained, and confidentiality was assured, but if they didn't want to be part of the study, residents could send in a card and they would be removed from the list. About 200 people returned the cards and opted out of the study (Wilson Hughes, personal communication). (And see Hughes [1984] for a detailed review of the methodology of the Garbage Project.)

By studying the detritus of ordinary people, researchers on the Garbage Project have learned interesting things about food consumption and waste among Americans. Squash is the favored baby food among Hispanics in the United States, and 35% of all food from chicken take-out restaurants is thrown away (Rathje 1992). You can accurately estimate the population of an area by weighing only the plastic trash. Children, it turns out, generate as much plastic trash as adults do (Edmondson 1988).

Early in the Garbage Project, researchers expected that people would not waste much beef during a shortage, but exactly the opposite happened in 1973. Two things were shown to be responsible for this finding.

First, as the shortage took hold, the price of beef rose, and people started buying cheaper cuts. Some residents did not know how to prepare those cuts properly, and this created more waste; others found that they didn't like the cheaper cuts and threw out more than they usually would have; and cheaper cuts have more waste fat to throw out to begin with.

Second, as the price continued to rise, people started buying greater quantities of beef, perhaps as a hedge against further price hikes. Inevitably, some of the increased purchases spoiled from lack of proper storage (Rathje 1984:17).

Rathje found the same pattern of consumer behavior during the sugar shortage of 1975. He reasoned that whenever people changed their food-buying and -consuming habits drastically, there would be at least a short-term increase in food loss. Conversely, when people use foods and ingredients that are familiar to them, they waste less in both preparation and consumption.

This led Rathje to compare the food loss rate among Mexican Americans and Anglos in Tucson and Milwaukee. “The final results of Mexican-American cooking,” Rathje said, “can be extremely varied—chimichangas, burros, enchiladas, tacos, and more—but the basic set of ingredients are very few compared to standard Anglo fare. Thus, Mexican-American households should throw out less food than Anglo households” (Rathje 1984:17–18). In fact, this is exactly what Rathje found in both Tucson and Milwaukee.

Beside Tucson and Milwaukee, studies of fresh household garbage have been done in New Orleans, Marin County (California), Mexico City, and Sydney (Australia).

Pros and Cons of Trace Studies

The most important advantage of trace studies is that they are nonreactive, so long as the people you are studying are kept in the dark about what you are doing. What happens when people are told that their garbage is being monitored? Rittenbaugh and Harrison (1984) compared data from an experimental group (people who were told that their garbage was being monitored) and a control group (people who were not told). There was no difference in the refuse disposal behavior of the experimental and control groups—with one important exception. The number of empty bottles of alcoholic drinks that showed up was significantly lower when people knew that their garbage was being monitored.

Where did the extra bottles go? Buried in the back yard? Stuffed in the trash cans of neighbors who were not in the sample? It remains a mystery.

In addition to being nonreactive, behavioral trace studies yield enormous amounts of data that can be standardized, quantified, and compared across groups and over time (Rathje 1979). Moreover, traces reflect many behaviors more accurately than informant reports of those behaviors.

In 1986, as part of a contract with the Heinz Corporation, Rathje and his colleagues asked women in the Tucson area if they had used any store-bought baby food in the past week. Uniformly, the Hispanic mothers insisted that they had not used any such product. You can guess the rest: They had as many baby-food jars in their garbage as did the Anglo households. As it turned out, 45% of the Hispanic women in Tucson at the time were working outside the home (Rathje and Murphy 1992). Those women were caught in a bind: They didn’t have time to prepare homemade baby food, but they couldn’t admit this to a stranger asking questions.

Trace studies like the Garbage Project have plenty of problems, however. Early in the project, it became apparent that garbage disposals were going to be a serious problem. The researchers constructed a subsample of 32 house-

holds, some of which had disposals, some of which did not. They studied these 32 households for 5 weeks and developed a “garbage disposal correction factor” (Rathje 1984:16).

As the project went on, researchers learned that some families were recycling all their aluminum cans, while others were throwing theirs in the trash. This made it difficult to compare households regarding their consumption of soft drinks and beer. Some families had compost heaps that they used as fertilizer for their vegetable gardens. This distorted the refuse count for those families. Garbage Project researchers had to develop correction factors for all of these biases, too (see Harrison 1976).

As with much unobtrusive research, the Garbage Project raised some difficult ethical problems. To protect the privacy of the households in the study, no addresses or names of household members are recorded. All personal items, such as photographs and letters, are thrown out without being examined. The hundreds of student sorters who have worked on the project have signed pledges not to save anything from the refuse they examine. All the sampling, sorting, and data analysis procedures are approved by the Human Subjects Research Committee of the University of Arizona.

The Garbage Project received consistent coverage in the press, both nationally and locally in Tucson. In 1984, after 10 years of work, Hughes reported that “no public concern over the issue of personal privacy has been expressed, and community response has been supportive” (Hughes 1984:42). With proper safeguards, trace measures can be used to generate useful data about human behavior.

Archival Research

One of the great advantages to doing archival research is that it is truly non-reactive. Whether you’re studying archival records of births, migrations, visits to a hospital, or purchases of hybrid seed, people can’t change their behavior after the fact. Of course, the original data might have been collected reactively, but that’s one reason why historians demand such critical examination of sources.

Another advantage of doing what Caroline Brettell calls “fieldwork in the archives” (1998) is that you can study things using archival data that would be too politically “hot” to study any other way. And archival research is inexpensive. Be on the lookout for interesting archival materials: government reports, personal diaries or photo collections, industrial data, medical records, school records, wills, deeds, records of court cases, tax rolls, and land-holding records.

Cultural Processes

Archival resources can be particularly useful in studying cultural processes through time. June Helm (1980) found that between 1829 and 1891, traders at the Hudson's Bay Company posts of the upper Mackenzie Delta had surveyed the Indians who traded at their stores. On the basis of those data, Helm concluded that, before 1850, the Indians of the area had practiced female infanticide. After 1850, missionaries were successful in stopping infanticide. Nancy Howell (1981), a demographer, subjected Helm's data to a sophisticated statistical analysis and corroborated Helm's conclusion.

Daniel Swan and Gregory Campbell (1989) studied the population records of 1877 to 1907 for the Osage reserve. They were able to show that from 1877 to 1887 the full bloods declined at 6.4% a year, while the mixed bloods increased at 7.3% a year. This had great consequences for the Osage because the full bloods and mixed bloods had formed voting blocs on economic issues. In particular, the full bloods resisted turning the reserve land into private property. Whites who married into the tribe fraudulently claimed tribal mixed-blood status. The mixed bloods were in favor of the private property measures.

Using fashion magazines going back to 1844, Alfred Kroeber made eight separate measurements of women's clothing in the United States and France (Kroeber 1919). He measured things like the diameter of the skirt at the hem, the diameter of the waist, the depth of décolletage (measured from the mouth to the middle of the corsage edge in front), and so on. After analyzing the data, Kroeber claimed to have found "an underlying pulsation in the width of civilized women's skirts, which is symmetrical and extends in its up and down beat over a full century; and an analogous rhythm in skirt length, but with a period of only about a third the duration" (*ibid.*:257). Kroeber offered his finding as evidence for long-cycle behavior in civilization.

Allport and Hartman (1931) criticized Kroeber for having been insufficiently critical of his sources. They found, for example, that the range in width of skirts for one year, 1886, was greater than the range Kroeber reported for 1859–1864 and that some years had very few cases on which to base measurements. If the data are suspect, Allport and Hartman concluded, then so are the regularities Kroeber claimed to have found (*ibid.*:342–343).

Richardson scoured the archives of fashion and extended the database from 1605–1936 (Richardson and Kroeber 1940). Before making measurements for all the new years included in the study, Richardson redid Kroeber's measurements for 1844–1846 and for 1919 and assured herself that she was coding each plate the same way Kroeber had done in 1919 (*ibid.*).

Lowe and Lowe (1982) reanalyzed the Richardson-Kroeber data for the 150 years from 1787–1936, using all the firepower of modern statistics and com-

puters. You'll be pleased to know that Kroeber's first analysis was vindicated: Stylistic change in women's dress is in stable equilibrium (changing with patterned regularity), and is driven by "inertia, cultural continuity, a rule system of aesthetic proportions, and an inherently unpredictable element" (ibid.:521).

Mulcahy and Herbert (1990) added data for the years 1937–1982 and found more variability in those 46 years than in the 150 years prior to 1937. For example, a plot of the moving average for skirt width from 1811–1926 has the same shape as the plot for 1926–1976. In other words, the cycle of skirt length had been cut by more than half in 1976.

The Problem with Archival Data

Kroeber's early work is being vindicated, but Allport and Hartman's critique was right on target in 1931. You can't be too critical of your sources. Archival data may appear clean, especially if they come packaged on computer disks and are coded and ready to be analyzed. In fact, they may be riddled with error. Consider carefully all the possible sources of bias (informant error, observer error, etc.) that might have been at work in the setting down of the data. Ask how, why, and under what conditions a particular set of archival data was collected. Ask who collected the data and what biases she or he might have had.

No data are free of error. In some parts of Mexico, the number of consensual unions is greater than the number of formal marriages, making court records about marriages problematic. In the United States, on the other hand, crime statistics are notoriously untrustworthy. Many crimes go unreported, and those that are reported may not be recorded at all, or may be recorded in the wrong category. In some countries, rural people may wait as long as 6 months to report a birth, and a significant fraction of their children may die within that period. (See Naroll [1962] and Handlin [1979] for discussions of data quality control in archival research.) It is almost always better to understand distortion in data than to throw them out. (For more on archival research in anthropology, see Brettell 1998.)

16



Introduction to Qualitative and Quantitative Analysis

This is the first of six chapters about analyzing data. In this chapter, we begin with the basics—what analysis is; how to store and manage data; and how to use matrices, tables, and flow charts to present the results of data analysis. Then we move on, in chapters 17 and 18 to methods for analyzing qualitative data, and in chapters 19, 20, and 21 to methods for analyzing quantitative data.

Qualitative, Quantitative

By a quirk of English grammar, the phrase “qualitative data analysis” is delightfully ambiguous. Unless someone spells it out in lots of words, you never know if the phrase means “the qualitative analysis of data” or the “analysis of qualitative data.” And, of course, the same goes for “quantitative data analysis.” Table 16.1 lays out the possibilities.

Cell *a* is the *qualitative analysis of qualitative data*. Interpretive studies of texts are of this kind. You focus on and name themes in texts. You tell the story, as you see it, of how the themes are related to one another and how characteristics of the speaker or speakers account for the existence of certain themes and the absence of others. You may deconstruct the text, look for hidden subtexts, and, in general, try to let your audience know—using the power of good rhetoric—the deeper meaning or the multiple meanings of the text.

Looking diagonally from cell *a*, cell *d* refers to numerical or statistical analysis of numerical data. Lots of useful data about human behavior come to us

TABLE 16.1
Qualitative/Quantitative Data Analysis

		<i>Data</i>	
		<i>Qualitative</i>	<i>Quantitative</i>
Analysis	Qualitative	a	b
	Quantitative	c	d

as numbers. Direct observation of behavior, village censuses, time allocation studies, closed-ended questions in surveys—all produce numerical data.

Cell *b* is the qualitative analysis of quantitative data. This can involve the search for patterns using visualization methods, like multidimensional scaling and hierarchical clustering. (We'll get to these methods in chapter 21.) Cell *b* is also about the search for, and the presentation of, *meaning* in the results of quantitative data processing. It's what quantitative analysts do after they get through doing the work in cell *d*. Without the work in cell *b*, cell *d* studies are sterile and superficial.

Which leaves cell *c*, the quantitative analysis of qualitative data. This involves turning the data from words or images into numbers. Scholars in communications, for example, tag a set of television ads from Mexico and the United States to test whether consumers are portrayed as older in one country than in the other. Political scientists code the rhetoric of a presidential debate to look for patterns and predictors of policies. Archeologists code a set of artifacts to produce emergent categories or styles, or to test whether some intrusive artifacts can be traced to a source.

Most quantitative analysis in the social sciences involves reducing people (as observed directly or through their texts) to numbers; most qualitative analysis involves reducing people to words—*your* words about the meaning of *their* words or actions or artifacts. I say “most” because a lot of analysis these days, qualitative and quantitative, involves visualization of data: not just looking for patterns in data, but showing the patterns as maps, networks, and matrices.

It's pretty obvious, I think, that each kind of data—qualitative and quantitative—and each kind of data reduction—qualitative and quantitative—is useful for answering certain kinds of questions. Skilled researchers can do it all.

What's Analysis?

Analysis is the search for patterns in data and for ideas that help explain why those patterns are there in the first place. The way I see it, analysis is

ultimately all qualitative. It starts before you collect data—you have to have some ideas about what you're going to study—and it continues throughout any research effort. As you develop ideas, you test them against your observations: Your observations may then modify your ideas, which then need to be tested again, and so on. Don't look for closure in the process. If you're doing it right, it never stops.

Most methods for quantitative analysis—things like factor analysis, cluster analysis, regression analysis, and so on—are really methods for data *processing* and for finding patterns in data. Interpreting those patterns is up to you. Interpretation—telling us what findings mean, linking your findings to the findings of other research—starts with ideas in your head and comes out in words on paper. It's a pretty qualitative exercise.

Don't worry about getting ideas. Once you have data in your hands, words or numbers, your hardest job will be to sort through all the ideas you get and decide which ones to test. And don't worry about seeing patterns in your data or about not being able to come up with causal explanations for things you see. It can happen very fast, often in a matter of hours or days after starting any research project, so be suspicious of your pet ideas and continually check yourself to make sure you're not inventing or at least not embellishing patterns.

Seeing patterns that aren't there happens all the time in research, qualitative or quantitative, just from eagerness and observer expectations. If you are highly self-critical, your tendency to see patterns everywhere will diminish as the research progresses.

The Constant Validity Check

The problem can also get worse, though, if you accept uncritically the folk analyses of articulate or prestigious informants. It's important to seek the emic perspective and to document folk analyses (Lofland 1971), but as your interviews and field notes pile up, try consciously to switch back and forth between the emic and etic perspectives. Check yourself from either buying into folk explanations or rejecting them without considering their possible validity. It's not hard to do this, it's just hard to remember to do it systematically. Here are some guidelines.

1. If you are interviewing people, look for consistencies and inconsistencies among knowledgeable informants and find out why those informants disagree about important things.
2. Whenever possible, check people's reports of behavior or of environmental conditions against more objective evidence. If you were a journalist and submitted a story based on informants' reports without checking the facts, you'd never get it

past your editor's desk. Why not hold anthropologists to the standard that journalists face every day?

3. Be open to negative evidence rather than annoyed when it pops up. When you run into a case that doesn't fit your theory, ask yourself whether it's the result of: (a) normal **intracultural variation**, (b) your lack of knowledge about the range of appropriate behavior, or (c) a genuinely unusual case.
4. As you come to understand how something works, seek out alternative explanations from key informants and from colleagues, and listen to them carefully. American folk culture, for example, holds that women left home for the workforce because of what are widely called "feminism" and "women's liberation." That's a popular emic explanation. An alternative explanation is that feminist values and orientations are supported, if not caused, by women being *driven* out of their homes and into the workforce by the hyperinflation during the 1970s that drove down the purchasing power of their husbands' incomes (Margolis 1984). Both the emic, folk explanation and the etic explanation are interesting for different reasons.
5. Try to fit extreme cases into your theory, and if the cases won't fit, don't be too quick to throw them out. It is always easier to throw out cases than it is to reexamine your own ideas, but the easy way out is hardly ever the right way in research.

Data Matrices

One of the most important concepts in all data analysis—whether we're working with quantitative or qualitative data—is the **data matrix**. There are two basic kinds of data matrices: **profile matrices** and **proximity matrices**. Figure 16.1 shows what these two kinds of matrices look like:

Profile Matrices

Across the social sciences, most data analysis is about *how properties of things are related to one another*. We ask, for example, "Is the ability to hunt related to the number of wives a man has?" "Is having an independent source of income related to women's total fertility?" "Are remittances from labor migrants related to the achievement in school of children left behind?" "Is the per capita gross national product of a nation related to the probability that it will go to war with its neighbors?"

This is called **profile analysis**. You start with series of *things*—units of analysis—and you measure a series of *variables* for each of those things. This produces a profile matrix, or, simply, a data matrix. *A data matrix is a table of cases and their associated variables*. Each unit of analysis is *profiled* by a

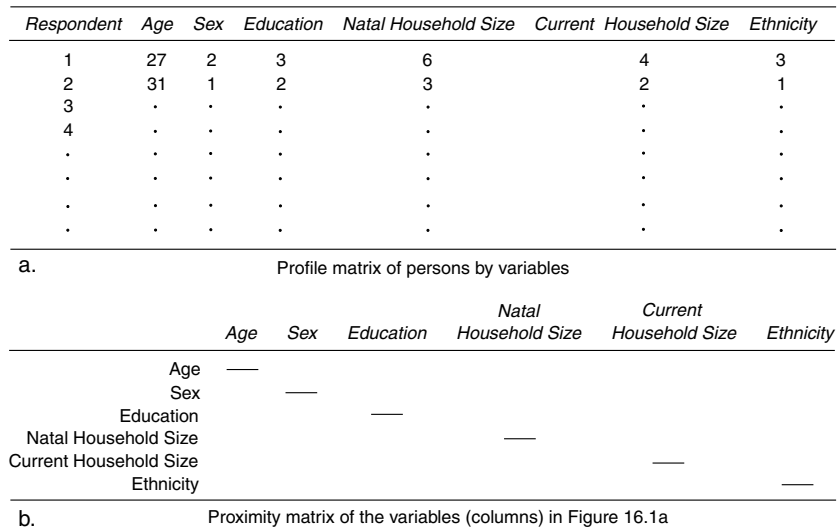


Figure 16.1. Two kinds of matrices: Profiles and proximities

particular set of measurements on some variables. Figure 16.1a shows the shape of a profile matrix.

Suppose you interview 100 women who are leaving an open-air village market in Cameroon. You record eight pieces of information about each person: (1) how much they say they spent on vegetables; (2) how much they say they’ve spent on meat or fish or poultry; (3) how much they say they’ve spent on other foods; (4) their age; (5) how many years they went to school; (6) whether they are married; (7) how many people are living in their home; and (8) how many children under the age of 15 are living in their home.

These data produce a 100 × 8 (read: 100-by-8) profile matrix. Each of the 100 cases (cases are people here) in your data file is characterized, or profiled by those six pieces of information. Table 16.2 shows the profiles for two cases:

Respondent #1 reported spending 330 CFA (Cameroonian francs) on vegetables, 420 CFA on meat, etc., and 215 CFA on other foods. She is 34 years old, went to school for 8 years, and is married. There are four people in her household, two of whom are children under the age of 15.

The units of analysis in a profile matrix are often respondents to a questionnaire, but they can be dowry records, folktales, churches—even time periods (1980, 1981, 1982, . . . 1991). You can also create a profile matrix from a set of coded texts or artifacts. (More about this in chapter 17.)

TABLE 16.2
Profile Matrix for Two Cases and Six Variables

Respondent	Variable							
	Spent on vegetables	Spent on meat, fish, or poultry	Spent on other foods	Age	Years in school	Married	How many in house	How many under 15 in house
1	330	420	215	34	8	1	4	2
2	195	250	300	60	6	0	7	3

Proximity Matrices

Profile matrices contain *measurements of variables for a set of items*. Proximity matrices contain *measurements of relations, or proximities, between items*. If the measurements in a proximity matrix tell *how close things are to each other* then you have a **similarity matrix**. If the measurements in a proximity matrix tell *how far apart things are from each other*, then you have a **dissimilarity matrix**.

If you've ever read one of those tables of distances between cities that you see on road maps, you've had experience with a dissimilarity matrix. The bigger the number in the cells, the more "dissimilar" two cities are. In this case, dissimilarity means that the larger the number in any cell, the further apart two cities are on the map.

If you've had a course in statistics and seen a correlation matrix, then you've had experience with a similarity matrix. The bigger the number in each cell—the higher the correlation—the more alike two things are.

Figure 16.1(b) shows a similarity matrix of variables. Imagine the list of variable names stretching several feet to the right, off the right-hand margin of the page, and several feet down, off the lower margin. That is what would happen if you had, say, 100 variables about each of your respondents. For each and every pair of variables in the matrix of data, you could ask: Are these variables related?

We'll need the concept of a proximity matrix when we get to multivariate analysis in chapter 21.

Presenting Results in Matrices and Tables

An important part of all analysis, qualitative and quantitative, is the production of visual displays. Laying out your data in table or matrix form and drawing your theories out in the form of a flow chart or map helps you understand

what you have and helps you communicate your ideas to others (Miles and Huberman 1994). Learning to build and use qualitative data matrices and flow charts requires practice, but you can get started by studying examples published in research journals.

Van Maanen et al. (1982), for example, compared a traditional commercial fishing operation in Gloucester, Massachusetts, with a modern operation in Bristol Bay, Alaska. Table 16.3 shows what they found in the analysis of their

TABLE 16.3
Contemporary Forms of Commercial Fishing

	<i>Traditional fishing</i> (e.g., Gloucester, MA)	<i>Modern fishing</i> (e.g., Bristol Bay, AK)
<i>Social Organization</i>		
backgrounds of fishermen	homogeneous	heterogeneous
ties among fishermen	multiple	single
boundaries to entry	social	economic
number of participants	stable	variable
social uncertainty	low	high
relations with competitors	collegial and individualistic	antagonistic and categorical
relations with port	permanent, with ties to community	temporary, with no local ties
mobility	low	high
relations to fishing	expressive (fishing as lifestyle)	instrumental (fishing as job)
orientation to work	long-term, optimizing (survival)	short-term, maximizing (seasonal)
tolerance for diversity	low	high
nature of disputes	intra-occupational	trans-occupational
<i>Economic Organization</i>		
relations of boats to buyers	personalized (long-term, informal)	contractual (short-term, formal)
information exchange	restrictive and private	open and public
economic uncertainty	low (long-term)	high (long-term)
capital investment range	small	large
profit margins	low	high
rate of innovation	low	high
specialization	low	high
regulatory mechanisms	informal and few	formal and many
stance toward authority	combative	compliant

SOURCE: J. Van Maanen et al., "An Occupation in Transition: Traditional and Modern Forms of Commercial Fishing," *Work and Occupations*, Vol. 9, pp. 193–216. © 1982, Reprinted by permission of Sage Publications.

qualitative field notes. Simple inspection of table 16.3 gives you an immediate feel for the results of Van Maanen et al.'s descriptive analysis.

The social organization of the traditional fishing operation is more homogeneous, more expressive, and more collegial than that of the modern operation, but profits are lower. Based on the qualitative analysis, Van Maanen et al. were able to state some general, theoretical hypotheses regarding the weakening of personal relations in technology-based fishing operations. This is the kind of general proposition that can be tested by using fishing operations as units of analysis and their technologies as the independent variable.

Donna Birdwell-Pheasant (1984) wanted to understand how differences in interpersonal relations change over time in the village of Chunox, Belize. She questioned 216 people about their relations with members of their families over the years, and simulated a longitudinal study with data from a cross-sectional sample. She checked the retrospective data with other information gathered by questionnaires, direct observations, and semistructured interviews. Table 16.4 shows the analytic framework that emerged from Birdwell-Pheasant's work.

Birdwell-Pheasant identified five kinds of relations: absent, attenuated, coordinate, subordinate, and superordinate. These represent the rows of the matrix in table 16.4. The columns in the matrix are the four major types of family relations: ascending generation (parents, aunts, uncles, etc.), siblings, spouse, and descending generation (children, nephews, and nieces, etc.).

Birdwell-Pheasant then went through her data and "examined all the available data on Juana Fulana and decided whether, in 1971, she had a coordinate or subordinate relationship with her mother (e.g., did she have her own kitchen? her own wash house?)." (In Latin America, Juan Fulano and Juana Fulana are the male and female equivalents of "so-and-so"—as in "Is so-and-so married?")

Birdwell-Pheasant repeated the process, for *each* of her 216 informants, for *each* of the four relations in table 16.4, and for *each* of the years 1965, 1971, 1973, 1975, and 1977. This required $216(4)(5) = 4,320$ decisions. Birdwell-Pheasant didn't have data on all possible informant-by-year-by-relation combinations, but, by the time she was through, she had a database of 742 "power readings" of family relations over time and was able to make some very strong statements about patterns of domestic structure over time in Chunox. This is an excellent example of the use of qualitative data to develop a theory, and the conversion of qualitative data to a set of numbers for testing that theory.

Stephen Fjellman and Hugh Gladwin (1985) studied the family histories of Haitian migrants to the United States. Fjellman and Gladwin found an elegant way to present a lot of information about those histories in a simple chart. Table 16.5 shows one chart for a family of four people in 1982.

This Haitian American family began in 1968 when Jeanne's father sent her to Brooklyn, New York, to go to high school. The single plus sign for 1968

TABLE 16.4
Birdwell-Pheasant's Matrix of Criteria for Assigning Values
to Major Relationships between People in Her Study

<i>Major Types of Relationships</i>				
<i>Values of Relationships</i>	<i>Ascending Generation</i>	<i>Siblings</i>	<i>Spouse</i>	<i>Descending Generation</i>
Absent	parents deceased, migrated permanently, or estranged	only child; siblings deceased, migrated permanently, or estranged	single or widowed; spouse migrated or permanently estranged	no mature offspring; all offspring deceased, or migrated permanently, or estranged
Attenuated	does not live with parents or participate in work group with parent; does visit and/or exchange food	does not live with siblings or participate in work groups with them; does visit and/or exchange food	separation, but without final termination of union; e.g., temporary migration	offspring do not live with parents or participate in work group with them; do visit and/or exchange food
Coordinate	participates in work group with parents, sharing decision-making authority	participates in work group with siblings under parents' authority; or works with siblings only, sharing decision making	married; in charge of own sex-specific domain with minimal interference from partner	participates in a work group with offspring, sharing decision-making authority
Subordinate	participates in work group with parent; parent makes decisions	participates in work group of siblings; other sibling(s) make decisions	individual's normal control within sex-specific domain is interfered with by spouse	dependent, elderly parent, unable to work
Superordinate	makes decisions for dependent, elderly parent who is unable to work	participates in work group with siblings; makes decisions for group	interferes with spouse's normal controls within sex-specific domain	heads work group that includes one or more mature offspring; makes decisions for group

SOURCE: "Personal Power Careers and the Development of Domestic Structure in a Small Community" by D. Birdwell-Pheasant, 1984, *American Ethnologist* 11(4). Reproduced by permission of the American Anthropological Association. Not for further reproduction.

TABLE 16.5
Family History of Haitian Migrants to Miami

<i>Year</i>	<i>Jeanne</i>	<i>Anna</i> <i>(mother)</i>	<i>Lucie</i> <i>(sister)</i>	<i>Charles</i> <i>(brother)</i>	<i>Marc</i> <i>(adopted</i> <i>son)</i>	<i>Helen</i> <i>(aunt)</i>	<i>Hughes &</i> <i>Valerie</i> <i>(cousins)</i>	<i>Number in</i> <i>Household</i>
1968	+							1
1971	+	+	+	+				4
1975	+	+	+	+	+			5
1976	+	+	-	-	+			3
1978	+	+	-	+	+		*	4
1979	+	+	-	+	+	+	*	5
1982	+	+	-	-	+	-	*	4

SOURCE: S. M. Fjellman and H. Gladwin, "Haitian Family Patterns of Migration to South Florida" *Human Organization*, Vol. 44, pp. 307, 1985. Reprinted with permission from the Society for Applied Anthropology.

shows the founding of the family by Jeanne. Jeanne's father died in 1971, and her mother, sister, and brother joined her in New York. Jeanne adopted Marc in 1975, and in 1976 she and her mother moved with Marc to Miami. Lucie and Charles remained together in New York. The two minus signs in the row for 1976 indicate that Jeanne's sister and brother were no longer part of the household founded by Jeanne.

Two years later, in 1978, Lucie got married and Charles joined Jeanne's household in Miami. Also in 1978, Jeanne began saving money and applying for visas to bring her cousins Hughes and Valerie to Miami. The asterisks show that these two people are in the process of joining the household. In 1979, Anna's sister, Helen joined the family, and in 1982 Charles went back to New York to live again with Lucie.

There is a lot of information in this chart, but the detail is gone. We don't know *why* Jeanne went to the United States in 1968; we don't know *why* Charles left Jeanne's household in 1976 or *why* he rejoined the group in 1978. Fjellman and Gladwin present seven of these family history charts in their article and they provide the historical detail in vignettes below each chart. Their purpose in reducing all the historical detail to a set of pluses and minuses, however, is to allow us to see the *patterns* of family growth, development, and decay.

Presenting Results: Causal Flow Charts

Causal maps represent theories about how things work. They are visual representations of ideas that emerge from studying data, seeing patterns, and

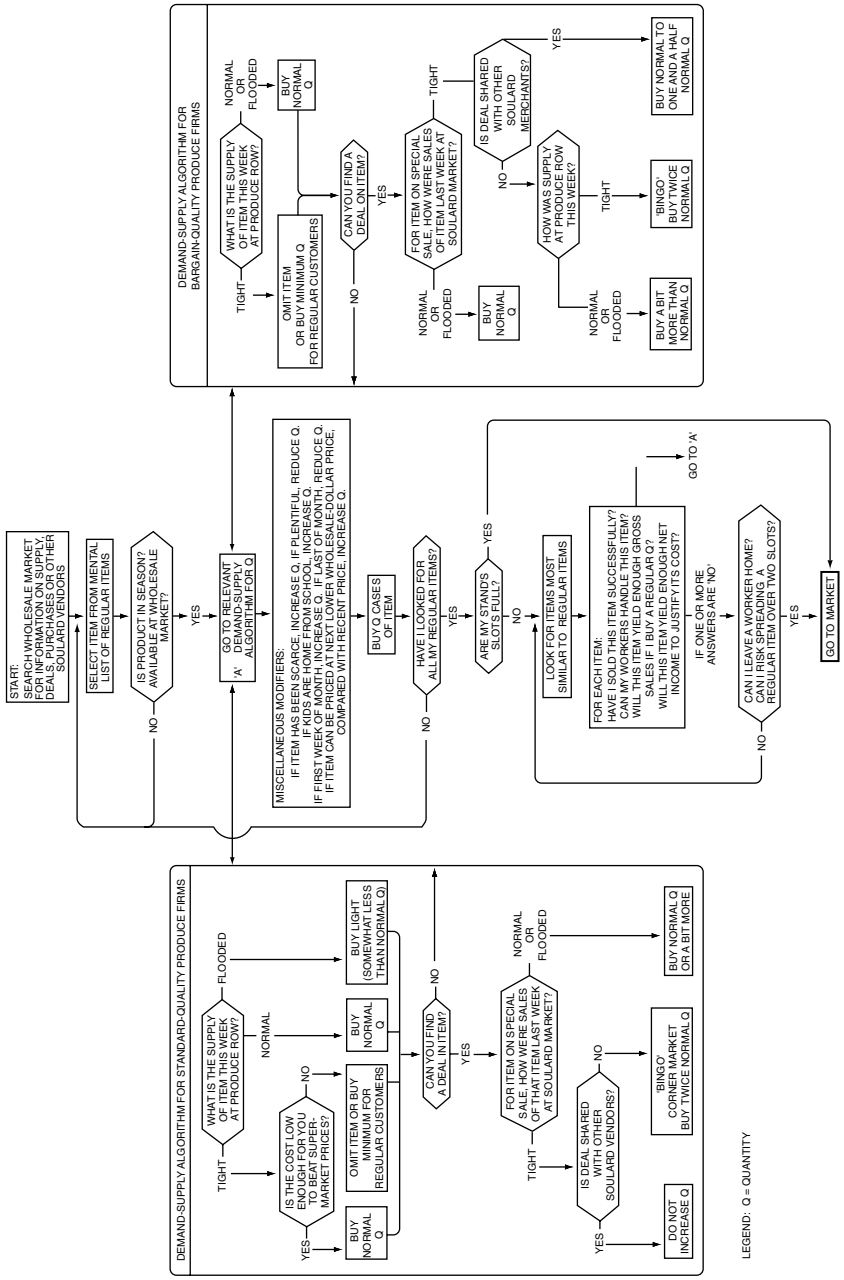


Figure 16.2. Plattner's model for how merchants in the Soulard Market in St. Louis decide what and how much produce to buy.
 SOURCE: S. Plattner, "Economic Decision Making in a Public Marketplace." *American Ethnologist* Vol. 9, p. 404, 1982. Reproduced by permission of the American Anthropological Association Not for further reproduction.

coming to conclusions about what causes what. Causal maps do not have to have numbers attached to them, although that is where causal modeling eventually leads. After all, it is better to know *how much* one thing causes another than to know simply that one thing *does* cause another. With or without numbers, though, causal models are best expressed as a **flow chart**.

A causal flow chart consists of a set of boxes connected by a set of arrows. The boxes contain descriptions of states (like being the youngest child, or owning a tractor, or being Catholic, or feeling angry), and the arrows tell you how one state leads to another. The simplest causal map is a visual representation of the relation between two variables

$$A \rightarrow B$$

which reads: “A leads to or causes B.”

Of course, real life is usually much, much more complicated than that. Look at figure 16.2. It is Stuart Plattner’s **algorithm**, based on intensive interviews and participant observation at produce markets in St. Louis, for how merchants decide what stock to buy. *An algorithm is a set of ordered rules that tell you how to solve a problem*—like “find the average of a list of numbers,” or, in this case, “determine the decisions of produce merchants.” (The capital letter Q in figure 16.2 stands for “quantity.”)

Read the flow chart from top to bottom and left to right, following the arrows. At the beginning of each week, the merchants seek information on the supply and cost of produce items. After that, the algorithm gets complicated. Plattner notes that the model may seem “too complex to represent the decision process of plain folks at the marketplace.” However, Plattner says, the chart “still omits consideration of an enormous amount of knowledge pertaining to qualities of produce at various seasons from various shipping areas” (Plattner 1982:405).

More about all this in chapter 18 when we get to models and matrices. Now, on to the nuts and bolts of data analysis.



Qualitative Data Analysis I: Text Analysis

There is growing interest these days in the analysis of texts. Little wonder. Most of the recoverable information about human thought and behavior in complex societies is naturally occurring text: books, magazines, and newspapers, of course, but also diaries, property transactions, recipes, correspondence, song lyrics, and billboards.

And there's more—lots and lots more. *Artifacts* (clay pots, houses, toys, clothing, computers, furniture) and *images* (family photo albums and videos, slasher films, sitcoms) are texts. *Behaviors* (making beer, laying out a garden, dressing the dead for funerals) and *events* (church ceremonies, homecoming games, blind dates) are texts. All these things come to us raw, in qualitative form. We can study them raw or we can code them—turn them into variables—and study the relations among the variables. Both approaches produce insight and understanding.

What's more, whole corpora of texts are becoming available in machine-readable form, either via the Internet or on CD-ROM—all U.S. Supreme Court opinions, the great works of the major world religions, all of Shakespeare's writings, every surviving ancient Greek and Latin text, 155 years of the *New York Times*, to name just a few, and more are added all the time. Ethnographic texts on a culture-area probability sample of 60 societies around the world are available on line from the Human Relations Area Files (more on HRAF later).

Scholars of social change have long relied on longitudinal survey data—the Gallup Poll (continually collected data since 1935) and the General Social Survey (almost every year since 1972) are just two of the hundreds of data

sets available for time series analysis. But longitudinal qualitative data are also plentiful. For a window on American popular culture, for example, take a look at the themes dealt with in country music and in *Superman* comics over the years. Or look at sitcoms from the 1950s and those made today.

In the 1950s, for example, Lucille Ball created a furor when she got pregnant and dared to continue making episodes of the *I Love Lucy* show. Now think about any episode of *Will and Grace* or *Sex and the City*. Or scan some of the recent episodes of popular soap operas and compare them to episodes from 30 years ago. Today's sitcoms and soaps contain many more sexual innuendos.

How many more? If you really wanted to measure that, you could code two representative samples of sitcoms and soaps, one from the 1950s and another from the last 10 years, and compare the codes statistically. That's content analysis.

Interpretivists, on the other hand, might be more interested in understanding the meaning across time of concepts like "flirtation," "deceit," "betrayal," "sensuality," and "love," or the narrative mechanisms by which any of these concepts is displayed or responded to by various characters.

Text analysis is for positivists and interpretivists alike. Both kinds of text analysis are easier to do today, with inexpensive, easy-to-use software for managing, coding, and retrieving texts.

The Great Anthropology Text Tradition

We'll get to text *analysis* later, but first I want to tell you about the great tradition of text *collection* in anthropology. Anthropologists have been big producers and collectors of texts, right from the beginning of the discipline. In 1881, Adolph Bastian argued that, with indigenous cultures of the world disappearing fast, "our guiding principle . . . in anthropology, prehistory, or ethnology should be to collect everything" (Bastian 1881:217; cited in Bunzl 1996:48).

For Franz Boas, Bastian's young colleague at the Berlin Museum, that meant collecting texts, and more texts, about the indigenous cultures of the world, *in the native languages* of those cultures. Boas trained his students (Kroeber, Lowie, Sapir, etc.) to collect verbatim text in as many indigenous American languages as they could reach.

This was really tough to do before voice recorders were invented, but that first generation of American anthropologists persevered and made use of every new technology they could to produce texts and more texts. In 1936, Margaret Mead and Gregory Bateson hauled heavy movie cameras to Bali and, with

cinematographer Jane Belo, shot hundreds of hours of *cinéma vérité* of Balinese dances and rituals—a treasure trove of fundamental ethnographic data (Belo 1960).

Boas practiced what he preached. He collected huge volumes of texts in several American Indian and Eskimo languages (Boas 1917, 1928a, 1928b). In the summer of 1888, while doing fieldwork in British Columbia, Boas met George Hunt. Hunt's father was a fur trader from Scotland and his mother was Tlingit, but Hunt grew up in Fort Rupert and became fluent in Kwakwala, the language of the Kwakiutl. Hunt lived as a Kwakiutl, married a Kwakiutl woman, and participated fully in the ceremonial life of the tribe. After Boas returned to New York, he and Hunt corresponded for 5 years and then, in 1893, Boas brought Hunt and several other Kwakiutl to the World's Fair in Chicago. It was there that Boas trained Hunt to write Kwakwala phonetically.

Hunt returned to British Columbia and Boas returned to New York. But from then on, for 40 years, Boas kept in touch with Hunt by mail, asking Hunt for information about this or that aspect of Kwakiutl culture. By the time Hunt died in 1933, he had produced 5,650 pages of Kwakiutl text. Boas relied on this rich corpus of descriptive ethnography for most of his reports about Kwakiutl life (Rohner 1966; see Boas and Hunt 1902–1905, 1906; Boas 1910, 1935–1943). In fact, of Boas's 5,000 pages of published work, 4,000 pages are unannotated translations of Kwakiutl language texts (Berman 1996:216) and there are thousands more pages in the Boas archives (at the American Philosophical Society in Philadelphia) that remain unpublished, waiting to be analyzed by another generation of scholars.

Native Ethnography

This was the first of what would become a tradition called **native ethnography**. It was carried forward by several of Boas's students, including Edward Sapir, whose first job after completing his Ph.D. under Boas's direction was at the Canadian Geological Survey, in Ottawa. One of the ethnographers who worked under his direction there was Marius Barbeau. In 1914, Barbeau went to the Northwest Coast to work with the Tsimshian and hired William Beynon as his interpreter. Beynon was 26 at the time, and only recently arrived from Victoria, but he spoke fluent Tsimshian (his mother was Tsimshian and his father was Welsh) (Anderson and Halpin 2000:4).

Barbeau wrote to Sapir that his new informant could recite Tsimshian myths and write them down in English at the same time. Sapir advised Barbeau to teach Beynon to write Tsimshian, as Sapir had taught his own informant to write Nootka, so that Beynon could work even more independently. "There is," said Sapir, "no absolute reason why every bit of material

that one utilizes in his work should have been personally obtained” (Sapir’s letter to Barbeau, February 1, 1915; quoted in Anderson and Halpin 2000:5–6.) This was the start of Beynon’s career as a professional informant. By the time he died in 1958, Beynon had sent Barbeau 54 Canadian government-issue notebooks of 50 pages each, filled with Tsimshian texts and with observations of Tsimshian daily life and ceremonies (ibid.:7).

Beynon never published a monograph based on his ethnographic notebooks, but four of his notebooks, detailing the ceremonies in 1945 surrounding the raising of five totem poles, were rescued from the archives and published verbatim (Anderson and Halpin 2000).

Another of Boas’s students, Paul Radin, however, took the method of native ethnography further. Radin studied the language and culture of the Winnebago Indians and, following his mentor’s lead, trained Sam Blowsnake to write in Winnebago. Blowsnake started writing his autobiography. After one day, he came to Radin, worried about what white people might think when they read it (Blowsnake confesses in the narrative to murdering a Potawatami). Radin reassured him, and Blowsnake polished off the manuscript in 2 days. Relying on a bilingual Winnebago-English interpreter, Radin translated and published the manuscript in 1920 as *Crashing Thunder: The Autobiography of a Winnebago Indian* (Radin 1983 [1920]).

Radin’s insistence that he “in no way influenced [Crashing Thunder], either directly or indirectly in any way” seems disingenuous today. And including that part about murdering a Potawatami was poor professional judgment, by today’s standards. But *Crashing Thunder* is, in my view, a milestone in the history of anthropology. Since then, the method of native ethnography has been greatly extended.

Fadwa El Guindi, for example, worked in the 1960s and 1970s with Abel Hernández Jiménez, on a native ethnography of the Zapotec people of Oaxaca. One of El Guindi’s colleagues, Harvey Rosenbaum, taught Hernández linguistics while El Guindi taught him anthropology (El Guindi 2004:141). El Guindi used open-ended interviewing to get Hernández to discuss four rituals: a baptism, a child’s funeral, a wedding, and an adult’s funeral. After each discussion (which covered things like the ordering of events in each ritual, who participates, and the meaning of ritual objects), Hernández wrote up his description of one of the rituals, and El Guindi collected parallel data on her own to compare with those of her colleague (El Guindi and Hernández Jiménez 1986:25).

El Guindi (2004:chap. 4) observes that some classic work in visual anthropology should also be counted as native ethnography. From 1963 to 1965, for example, Asen Balicki worked with Itimanguerk, a Netsilik Inuit hunter, producing a monumental film record of indigenous Netsilik culture: nearly 50

miles of film, including nine finished films totaling nearly 11 hours (Balicki 1989; El Guindi 2004:129–34). And in 1966, Sol Worth and John Adair gave 16-mm movie cameras to seven Navajo men and women to learn about Navajo life “through Navajo eyes,” the title of the book describing this famous project (Worth and Adair 1970, 1972; El Guindi 2004:140–146).

Ignacio Bizarro Ujpán has been the long-term informant and friend of anthropologist James Sexton. Between 1972 and 1974, at Sexton’s urging, Bizarro wrote an autobiography in Spanish. He also began a diary in 1972, which he kept for 5 years. Sexton edited the diary, removing repetitious events and what he felt was trivial material detailing “the kinds of meals he [Bizarro] ate each day, the saying of his nightly prayers, his daily work routine, how well he slept, and common illnesses like colds and headaches” (Sexton 1981:4).

Sexton supplemented the diary with taped interviews. In his method, he inserts his questions in italics into Bizarro’s narrative, to show “what kind of information Ignacio volunteered and how much of it is material I solicited from my own perspective” (*ibid.*; see also Bizarro Ujpán [1985, 1992, and Sexton and Bizarro Ujpán 2001] for the continuation of the Bizarro diaries).

The Nāhñu Native Ethnography

My own contribution to this tradition of native ethnography has been in collaboration with Jesús Salinas Pedraza, a Nāhñu teacher from the Mezquital Valley in central Mexico. Jesús and I met in 1962 when I went to the Mezquital to study Nāhñu for my M.A. In 1971, I told him that I was thinking of writing an ethnography of the Nāhñu. He said that he might like to write one himself, and so for the next 16 years, he wrote, and I translated and annotated, a four-volume, 232,000-word book on Nāhñu culture in Nāhñu (Salinas Pedraza and Bernard 1978; Salinas Pedraza 1984; Bernard and Salinas Pedraza 1989).

Just as I had developed a writing system for Nāhñu and had taught Jesús to intellectualize about the phonology of his language, I supposed that I would teach him to take field notes about ceremonies and to write about all the details of everyday life. As I said in 1989, in the introduction to the Nāhñu ethnography:

I saw him writing up a bunch of raw facts and me guiding him in the analysis of those facts, as we teased out the underlying meaning of the texts produced by his unschooled efforts. I saw myself guiding Jesús, monitoring and supervising his work, and seeing to it that the ethnography he would produce would meet the high standards of modern anthropology. I wanted to be able to say that, as the

professional anthropologist of this team, I had laid academic hands on Jesús's work and that the result was much better for it. (Bernard and Salinas Pedraza 1989:24)

I was soon disabused of these illusions. Jesús Salinas wrote his 232,000-word ethnography of the Ñähñu with very little active coaching from me.

WOLCOTT'S CRITIQUE OF NATIVE ETHNOGRAPHY

Very little active coaching, but still. . . . Harry Wolcott (1999) offers two incisive critiques of the Ñähñu native ethnography project: (1) that I had more influence on Jesús Salinas's writing of the Ñähñu ethnography than I imagined in 1989, and (2) that the Ñähñu ethnography is not really an ethnography because Jesús Salinas is not really an ethnographer.

Wolcott is right about the first point. As he points out (*ibid.*:151), early on in the Ñähñu ethnography project, I presented Salinas with a Spanish-language version of the *Outline of Cultural Materials*, the detailed codebook developed by George Peter Murdock to categorize ethnographic materials on cultures of the world (Murdock 1971). This reflected my own training, and was hardly a totally hands-off approach to the content of Jesús's work.

In fact, my influence extended far beyond just advocating that Salinas pay attention to the categories of the OCM. By the time we completed the translation of the Ñähñu native ethnography in 1988, Jesús and I had known one another for 26 years. I had asked him countless questions about Ñähñu culture before we ever got started on the project and countless more as we sat, elbow to elbow, translating the ethnography from Ñähñu into English (we worked through Spanish as an intermediary language). I forced him to think out loud about things that he might otherwise never have contemplated consciously. In the end, the content of the Ñähñu native ethnography was the result of constant negotiation, constant interaction between Jesús Salinas and me.

On the other hand, Jesús was not chained to the OCM and he went far beyond simply addressing my questions. He wrote about things that I could not have asked about, simply because of my ignorance of Ñähñu culture. I made plain that I was interested in the uses of plants, but he wrote about plants that I didn't know existed.

I asked him to begin the Ñähñu ethnography with a chapter on the setting and explained that this was about the geography, the climate, the flora and fauna, and so on. Again, this was straight out of my own training, and Jesús followed my instructions, but the result was nothing like anything I *would* have written, because it was nothing I *could* have written. As the ethnography begins, Jesús walks the reader through the Mezquital Valley, naming the

mountains and describing the villages in great detail: which villages produce lime for tortillas, which produce charcoal, and so on. He asked me if he could include information on the weather and the seasons and the winds and what effect they all had on people, plants, and animals. I could not even have formulated the questions in Nāhñu to retrieve that kind of information.

Others may choose to pursue the method of native ethnography differently, but I would do it the same way again. I would provide a framework and a long list of topics (the OCM, for example). I wouldn't tell the native ethnographer what content to provide, but, as I did with Jesús, I would ask plenty of questions, based on my own knowledge of the local culture—and on the holes I recognize in that knowledge. I believe strongly that we need more, much more, of this kind of work in anthropology.

Wolcott's second critique—that the Nāhñu ethnography is not really an ethnography because Jesús Salinas is not really an ethnographer—raises a really interesting question: Who gets to be called an ethnographer? One school of thought is captured in Paul Bohannan's aphorism that "without an ethnographer there is no ethnography" (Bohannan 1995:157). Wolcott quotes this maxim to lead off his penetrating discussion of insider and outsider ethnography (Wolcott 1999:131). In this perspective, Jesús Salinas's work is "the stuff out of which ethnography is made," but it is not ethnography (*ibid.*:153). Just as it is the rare ethnographer who can master the details of a culture the way Salinas masters his, says Wolcott, "it would be an even rarer 'native' who would have an intuitive grasp of ethnography" (*ibid.*:155).

In my view, this makes far too much of interpretation and not nearly enough of description. Here is Salinas, the ethnographer, presenting a description of one insect—illustrating what Wolcott (*ibid.*:154) identifies, correctly, as crushing detail:

1163. The nigua. These are the same size and have the same characteristics as squirrel and cat fleas. The nigua also jumps and turns up everywhere. One can see them in the feet of pigs. The feet are cut open by the fleas which deposit their eggs between the toes. The eggs are in a large, white ball in the pig's flesh. 1164. This is cured by putting on criolina so that the foot of the animal suppurates and the flea egg-balls run out. This flea also affects people's feet; they go between the toes. At first, they just itch; though one knows it itches, one may think it's just any old flea. 1165. But after three days, the itching gets worse, and it begins to suppurate and it grows worse and worse. When it gets like this it is opened up with a needle and a ball comes out, filled with eggs. 1166. A hole remains where the ball is taken out, as if a piece of flesh were removed. One must burn the ball so that the fleas won't be born. It is said that if it thrown on the ground, the heat from the earth will hatch the eggs. 1167. So that they won't hatch, the balls are removed from people's feet and thrown into the fire to burn up. If one allows the

swelling on the feet to burst by itself, then it hurts even more. 1167. Niguas lay their eggs in the webbing between the toes. Of course, this happens to people who don't wear huaraches; niguas are an affliction of those who go barefoot. Sometimes it causes just itching and other times it itches and hurts. (Bernard and Salinas 1989:209)

And here is Salinas, the storyteller, talking about religion and playing, what Wolcott says is “the docent, giving us a carefully arranged tour” (Wolcott 1999:154):

1136. Right next to San Clemente is another small village called El Nogal. This community belongs to Orizabita and only has 17 full citizens. All together, among the men, women, and children, there are 55 people there. There is no church of either of the two religions. The people of San Clemente tried to impose Protestantism on the people of El Nogal, but the people of El Nogal did not like that religion. Because of that, and because of other things that they did to them, the people of El Nogal abandoned Protestantism, became even stronger believers in Catholicism, and opted to follow their pueblo, which is Orizabita. (Bernard and Salinas 1989:533)

There were, to be sure, complaints over the years that Boas's work was nothing more than a mountain of raw facts. Leslie White complained that Boas's Kwakiutl texts were not intelligible because they were without commentary (White 1963:55), and George Peter Murdock mocked Boas's “five-foot shelf” of monographs about the Kwakiutl as contributing little to understanding the social structure of the Kwakiutl (Murdock 1949:xiv, n. 5).

From my perspective, ethnography has not outlived the positivist, descriptive function it had for Boas and his students and the goal of native ethnography is the same today as it was in Boas's time: Get it down and get it right, in the natives' own languages. “Getting it right” is a goal, not a given, so we keep trying to improve the method. One thing we can do is to provide training (even more than I offered Jesús Salinas) in what scholars of culture are interested in knowing. Another is to provide extensive commentary and explanations of the native-language materials (as I did in the *Ñähñu* native ethnography [Bernard and Salinas Pedraza 1989]).

More important than commentary or explanation, however, is the ineluctable presence of those 232,000 words of *Ñähñu* that Salinas produced about *Ñähñu* culture. Many of the languages in which Boas's students collected ethnographic texts are no longer spoken, or will soon die out. The texts collected by the Boasians for 50 years are a precious resource for all scholars interested in the diversity of human culture. Debates about -isms—structuralism, materialism, postmodernism—come and go, but texts are forever.

Is an N of 1 Really Science?

Of course, Salinas's native ethnography is one person's report of selected aspects of Nāhñu culture. What if other Nāhñu don't agree with Salinas on key issues? What if other Nāhñu think that Salinas focused too much on the flora and fauna of the Mezquital Valley and ignored other elements of the culture? Well, all ethnographies are written by *someone*. Why shouldn't other Nāhñu disagree with Salinas about his take on key issues? "Why not," Jesús once asked me, "invite a group of Nāhñu to read the Nāhñu version of the ethnography, and offer new interpretations?" I can hardly wait.

Other Kinds of Native Ethnography

The term "native ethnography" covers a lot of ground. At one end of the spectrum are the works of George Hunt, Jesús Salinas, Crashing Thunder, Abel Hernández Jiménez, and Ignacio Bizarro Ujpán, written entirely in the nonliterary, indigenous language of the native ethnographer.

At the other end are the works of indigenous people who have become professional anthropologists and who write about their own culture in one of the major literary languages of the world, like English, Japanese, French, Spanish, etc. Included here are the works of scholars like Jomo Kenyatta (1938) and Victor Uchendu (1965), who wrote ethnographies of the Gikuyu (in Kenya) and Igbo (in Nigeria), respectively. Both were bilingual speakers of English and their tribal language; both were trained as anthropologists; and both wrote in English. Kenyatta studied with Bronislaw Malinowski at the London School of Economics in the 1930s. Uchendu's ethnography of the Igbo was written as his M.A. thesis in anthropology when he was a student at Northwestern University (Uchendu 1965).

Another genre comprises autobiographies by speakers of nonliterary languages who were not trained as ethnographers but who wrote in a major literary language. Between 1964 and 1968, Refugio Savala, a Yaqui poet, wrote his richly detailed autobiography in English (Savala 1980). In his narrative, Savala tries to describe Yaqui culture in general by showing his family's participation in it (Sands 1980).

Other examples of this genre for North American Indians include *The Middle Five: Indian Schoolboys of the Omaha Tribe* by Francis La Flesche (1963 [1900]) about life at a boarding school; *Wah'Kon-Tah: The Osage and the White Man's Road*, by John Joseph Mathews (1968 [1932]) about life under a particular Indian agent on the Osage Reservation in the late 1800s; and *The Way to Rainy Mountain* by N. Scott Momaday (1969), in which Momaday, who holds a Ph.D. in English from Stanford, tells of his attempt to seek his

identity as a Kiowa from his grandmother and from other elders with whom he consulted.

Yet another genre of native ethnography comprises the “as told to” autobiographies. One famous example is *Black Elk Speaks*. John Neihardt, an epic poet from Nebraska, met Black Elk, an Oglala Sioux, in 1930. A year later, Neihardt and his daughters went to Black Elk’s home on the Pine Ridge Reservation. Black Elk spoke in Sioux and his words were translated by his son Ben, who had studied at Carlisle University.

Neihardt says that it was his own function “to translate the old man’s story, not only in the factual sense—for it was not the facts that mattered most—but rather to recreate in English the mood and manner of the old man’s narrative” (Neihardt 1972 [1932]:xii).

Other examples of this genre for North American Indians include the as-told-to autobiography of Don Talayesva, a Hopi (Simmons 1942); of John Lame Deer, a Miniconjou Sioux (Lame Deer and Erdoes 1972); of Two Leggings, a Crow (Nabokov 1967); and of Left Handed, a Navajo (Dyk 1938; Dyk and Dyk 1980). Here is Ruth Dyk, telling us how the story of Left Handed was actually recorded. “Since Left Handed did not know English he told his story in Navajo, and it was translated bit by bit by Philip Davis, a Navajo. Left Handed would speak a minute or two, Philip would translate, and my husband would write down the translation” (Dyk and Dyk 1980:xvii).

Roger Keesing recorded the life story of ’Elota, a Big Man of the Kwaio in the Solomon Islands (Keesing 1978). The 15 recording sessions took place over a period of 4 months. Keesing translated the tapes and provides a sketch of Kwaio culture and history. Andrew Strathern recorded and translated the life history of Ongka, a Big Man of the Kawelka in Papua New Guinea (Strathern 1979). At first, Strathern prompted Ongka, but he soon left the Big Man alone with the tape recorder. “I would return to the room from time to time to change the cassette over,” reports Strathern, “and would find him fully engrossed, gesturing and smiling into the microphone” (ibid.:x).

During two field trips to Botswana in 1971 and 1975, Marjorie Shostak collected 30 hours of taped interviews with Nisa, a !Kung woman. Nisa was about 50 in 1971—an advanced age for a !Kung woman at that time. She spoke only in !Kung, and Shostak’s English translation ran to hundreds of typewritten pages (Shostak 1981:42–43). Shostak edited the tapes—in some cases bringing parts of stories together to create continuous chronological narrative—to produce Nisa’s moving autobiography.

Mountain Wolf Woman was the sister of Crashing Thunder. Her autobiography was recorded in Ann Arbor, Michigan, in 1958, at the home of Nancy Lurie. Over a period of 5 weeks, during which Mountain Wolf Woman lived with the Luries, she told her story in Winnebago, and Lurie recorded it on

tape. Then, using the Winnebago tapes as a guide, and a second tape recorder, Mountain Wolf Woman told her story again, in English (Lurie 1961).

Lurie and Mountain Wolf Woman translated the Winnebago tape together, and Mountain Wolf Woman's grand-niece, Frances Thundercloud Wentz, helped Lurie produce the final, detailed translation. Lurie is critically aware of her influence on the final product: "The final preparation of an acceptable English narrative from a literary and scholarly point of view required decisions for which I must take full responsibility." These decisions, said Lurie, include "choice of tenses, equivalents of idiomatic expressions, insertion of words necessary for clarification, and the like" (ibid.:95).

Lurie also recorded comments that Mountain Wolf Woman made during the 5-week stay and inserted some of these comments into the final published narrative. She had them set in italics to show where she had added text from materials other than the original tapes.

All of these different kinds of native ethnographies offer glimpses—some more descriptive, some more interpretive—into the lived experience of people in other cultures. But all of them depend first on the collection of mountains of texts.

Some Traditions of Text Analysis

Analyzing texts is another matter. There is no single method for doing it. Some of the major traditions of text analysis include **hermeneutics (interpretive analysis)**, **narrative and performance analysis**, **schema analysis**, **discourse analysis**, **grounded theory**, and **content analysis**.

Hermeneutics/Interpretive Analysis

Modern hermeneutics derives from biblical hermeneutics, also called **biblical exegesis**. (Recall from chapter 1 that the word "hermeneutics" comes from Hermes—he of the winged hat—Greek messenger of the gods, whose job it was to interpret the will of the gods to humans.) In the Western tradition of exegesis, the Old and New Testaments are assumed to contain eternal truths, put there by an omnipotent creator through some emissaries—prophets, writers of the Gospels, and the like. The idea is to continually interpret the words of those texts in order to understand their original meaning and their directives for living in the present.

Formal rules for biblical exegesis were developed by the early Talmudic scholars, about a hundred years after the death of Jesus of Nazareth. For exam-

ple, one of the rules was that “the meaning of a passage can be derived either from its context or from a statement later on in the same passage.” Another was that “when two verses appear to contradict one another, a third verse can be discovered which reconciles them.”

The 13 rules for interpreting scripture remains to this day part of the morning service among Orthodox Jews. In the 18th and 19th centuries, New Testament scholars used hermeneutic reasoning to solve what was called the “synoptic problem”—that is, to determine the order in which the four Gospels were written. (See Farmer [1976] and Stein [1987] for detailed histories of the problem.)

The method of close interpretation of text has been extended from biblical studies to other areas. Constitutional law in the United States is a form of hermeneutics. Jurists take it as their task to consider what the writers of each phrase in the U.S. Constitution meant at the time and to interpret that meaning in light of current circumstances. It is exegesis on the U.S. Constitution that has produced different interpretations across time about the legality of slavery, abortion, women’s right to vote, the government’s ability to tax income, and so on.

Extending Hermeneutics to All Kinds of Texts

The hermeneutic method has been extended to the study of all kinds of texts, including images, conversations, and even song lyrics. Michael Herzfeld (1977) studied renditions of the *khelidonisma*, or swallow song, sung in modern Greece as part of the welcoming of spring. Herzfeld collected texts of the song from ancient, medieval, and modern historical sources and recorded texts of current-day renditions in several locations across Greece. His purpose was to show that inconsistencies in the texts come not from “some putative irrationality in the processes of oral tradition” but are, in fact, reflections of structural principles that underlie the rite of passage for welcoming spring in rural Greece.

To make his point, Herzfeld looked for anomalies across renditions—like “March, my good March” in one song compared to “March, terrible March” in another. Herzfeld claims that the word “good” is used ironically in Greek where the referent is a source of anxiety.

Is March a subject of symbolic anxiety for Greek villagers? Yes, says, Herzfeld, it is, and we can tell that it is because of widely observed practices like avoidance of certain activities during the *drimata* (the first 3 days of March). Herzfeld supports his analysis by referring to the *drimes*, a word that denotes the first 3 days of August, which are associated with malevolent spirits. Since March is the transition from winter to summer and August is the transition

from summer to winter, Herzfeld concludes that there is symbolic danger associated with these mediating months. He finds support for this analysis in the fact that February is never referred to with an unequivocally good epithet.

This is hermeneutic analysis—the search for meanings and their interconnection in the expression of culture. The method for doing this kind of analysis requires deep involvement with the culture, including an intimate familiarity with the language, so that the symbolic referents emerge during the study of those expressions—as in the study of texts here. You can't see the connections among symbols if you don't know what the symbols are and what they are supposed to mean.

Narrative and Performance Analysis

In narrative analysis, the goal is to discover regularities in how people tell stories or give speeches. This is achieved mostly through the analysis of written text. Dell Hymes's work is instructive. In 1977, Hymes reported that “the narratives of the Chinookan peoples of Oregon and Washington can be shown to be organized in terms of lines, verses, stanzas, scenes, and what many call acts.” Hymes felt that this discovery might be relevant to many indigenous languages of the Americas (1977:431).

Chinookan is a family of American Indian languages from the northwest coast of North America. Shoalwater Chinook and Kathlamet Chinook are two mutually unintelligible languages (related in the way French and Spanish are both Romance languages), but Franz Boas had run into an informant who was fluent in both Shoalwater and Kathlamet and had collected texts in both languages between 1890 and 1894. Hymes examined those texts as well as texts from Clackamas Chinook (collected in 1930 and 1931 by Melville Jacobs) and in Wasco-Wishram Chinook (collected by Sapir in 1905, by Hymes in the 1950s, and by Michael Silverstein in the 1960s and 1970s).

What Hymes found was that features of Chinook that might have seemed idiosyncratic to the speakers of those three Chinook languages—Shoalwater, Kathlamet, and Clackamas Chinook—were actually “part of a common fabric of performance style,” so that the three languages “share a common form of poetic organization” (Hymes 1977:431).

This was a truly important discovery. It made clear once and for all that Native American texts have something to contribute to a general theory of poetics and literature. Hymes discovered the existence of verses, by recognizing repetition within a block of text. “Covariation between form and meaning,” said Hymes, “between units with a recurrent Chinookan pattern of narrative organization, is the key” (ibid.:438).

In some texts, Hymes found recurrent linguistic elements that made the task easy. Linguists who have worked with precisely recorded texts in Native American languages have noticed the recurrence of elements like “Now,” “Then,” “Now then,” and “Now again” at the beginning of sentences. These kinds of things often signal the separation of verses. The trick is to recognize them and the method is to look for “abstract features that co-occur with the use of initial particle pairs in the narratives” of other speakers who use initial particle pairs. The method, then, is a form of controlled comparison (ibid.:439).

In a series of articles and books (1976, 1977, 1980a, 1980b, 1981), Hymes showed that most Native American texts of narrative performance (going back to the early texts collected by Boas and his students and continuing in today’s narrative performance by American Indians as well) are organized into verses and stanzas that are aggregated into groups of either fives and threes or fours and twos. Boas and his students organized the narratives of American Indians into lines.

According to Virginia Hymes, this hid from view “a vast world of poetry waiting to be released by those of us with some knowledge of the languages” (1987:65). Dell Hymes’s method, according to Virginia Hymes, involves “working back and forth between content and form, between organization at the level of the whole narrative and at the level of the details of lines within a single verse or even words within a line” (ibid.:67–68). Gradually, an analysis emerges that reflects the analyst’s understanding of the larger narrative tradition and of the particular narrator.

This emergent analysis doesn’t happen miraculously. It is, Virginia Hymes reminds us, only through close work with many narratives by many narrators that you develop an understanding of the narrative devices that people use in a particular language and the many ways they use those little devices (ibid.).

But all this depends on having texts—lots of them, like those that Boas and his students left us—marked for features like voice quality, loudness, pausing, intonation, stress, and nonphonemic vowel length (Hymes 1977:452–53).

Dennis Tedlock (1987) showed the exegetical power that linguistic methods can bring to text analysis. He had translated the *Popol Vuh*, a 16th-century Quiché Maya manuscript that had been written out by Francisco Ximénez, a missionary of the time. The *Popol Vuh* is one of those big epics, like the *Iliad* or *Beowulf* that were meant to be recited aloud. Is it possible, Tedlock asked, to analyze the text and figure out how to narrate it today as performers would have done in ancient times?

In doing his translation of the *Popol Vuh*, Tedlock had relied on Andrés Xiloj, a modern speaker of Quiché. Xiloj had not been trained to read Maya, but he was literate in Spanish and made the transition very quickly. “When he

was given his first chance to look at the *Popol Vuh* text, he produced a pair of spectacles and began reading aloud, word by word” (ibid.:145).

Like many medieval manuscripts in Europe, Ximénez’s rendition of the *Popol Vuh* was more or less an undifferentiated mass of text with almost no punctuation. In other words, Tedlock had no clues about how a performer of the narrative 400 years ago might have varied his timing, emphasized this or that segment, used different intonations, and so forth.

Tedlock’s solution was to study stylized oral narratives (not just casual speech) of modern speakers of the Quiché. He recorded speeches, prayers, songs, and stories and looked for phrases and patterns in the wording that had analogs in the *Popol Vuh* (ibid.:147). He devised special punctuation symbols for marking pauses, accelerations, verse endings, and so on and applied them to the *Popol Vuh*. It’s in the use of those written marks that we see Tedlock’s analysis—his understanding of how a performance went.

Tedlock then made systematic comparison across other ancient texts to look for recurrent sound patterns that signify variations in meaning. (Think of how we use rising intonation at the end of sentences in English to signify a question, and how some people in our society use the same intonation in declarative sentences at the beginning of phone conversations when the object is to jar someone’s memory, as in: “Hi, this is Mary? I was in your intro class last semester?”) What Tedlock found was that Quiché verse has the same structure as ancient Middle Eastern texts—texts that, he points out, predate Homer. In fact, Tedlock concluded it is the same structure found in all living oral traditions that have not yet been influenced by writing (ibid.:146; and see Tedlock 1977).

Joel Sherzer (1994) presents a detailed analysis of a 2-hour performance that he recorded in 1970 of a traditional San Blas Kuna chant (the Kuna are an indigenous people from Panama). The chant was rendered by Chief Olopinikwa and recorded on tape. Like many linguistic anthropologists, Sherzer had taught an assistant, Alberto Campos, to use a phonetic transcription system. After the chant, Sherzer asked Campos to transcribe and translate the tape. Campos put Kuna and Spanish on left- and right-facing pages (ibid.:907).

Another chief responded to Olopinikwa, usually with some utterance like “So it is.” Sherzer noticed that Campos left those responses out of the translation, but it was just those responses that turned out to be markers for verse endings in the chant. And Campos left out words like “Thus” at the beginning of a verse and phrases like “it is said, so I pronounce” spoken by the narrator at the end of a verse. These phrases also contribute to the line and verse structure of the chant. Campos, in other words, edited the chant to give Sherzer the essentials, but it was the seemingly inessential things—those little framing

phrases—that defined cultural competence in performance of the narrative (ibid.:908).

A key method of text analysis in ethnopoetics is text presentation. As Sherzer discovered, verse breaks in Kuna narratives are determined by regular turn-taking between narrating chief and responding chief. In presenting Chief Olopinikwa's performance in print, Sherzer begins each line of the main narrative flush left and indents the lines of the responding chief.

A few years earlier, in his presentation of *The Hot Pepper Story*, Sherzer (1990:178) used a highly literal translation. In the original text, there are many repetitions of a small number of words and themes. Sherzer felt that a liberal translation would fail to capture the poetics of performance, so in his commentary he describes the thematic elements in the text but uses the device of literalness in the translation to draw the reader's attention to those elements.

Translations are texts in their own right and can be analyzed just as original texts can. William Hanks (1988) reviewed Munroe Edmonson's (1986) translation of *The Book of Chilam Balam of Chumayel*. Edmonson had translated and annotated the original Mayan corpus into 5,514 lines of text, changing the format of presentation in the process. In the original, the lines had run clear across the page, but Edmonson presented the text in short lines to emphasize what he considered to be the verse structure. Hanks analyzed not only the Mayan text but the literary style that Edmonson used in his presentation. This method, of course, is well known from generations of work on literary criticism applied to translations (Rose 1997) and on linguistics applied to translation. (See Nida and Taber [1969] and Nida [1975] for the classic work here.)

In translating Ñähñu (Otomí) parables, folktales, and jokes, Jesús Salinas and I presented a fully literal translation and a fully liberal translation, in addition to a transcription of the Ñähñu (Bernard and Salinas 1976). Figure 17.1 is a piece of the text from that effort:

At the time, I felt that I couldn't show some of the characteristics of the original, idiomatic Ñähñu in a free, idiomatic English translation, so I suggested to Salinas that we present both, along with a literal translation from Ñähñu to English. That way, readers could decide for themselves if my idiomatic translation from Ñähñu to English made sense. (The original publication, in fact, was marked for tone and for pauses, which I've left out here.) Later, in translating Salinas's four-volume ethnography of the Ñähñu, I didn't have the luxury of including either the Ñähñu or a literal translation, and so I settled on a middle course: The English is idiomatic and grammatical but I tried deliberately to reflect Salinas's style (see Bernard and Salinas 1989).

We're still experimenting with methods for presenting texts of indigenous performance that capture the subtleties of performance. We can't know for sure if a particular presentation actually captures the regular features of narrative, but the same could be said for any interpretive effort in which anthropol-

Ñähñu

1. n'a zi ñ'ohō rc zi memapa šā maa rā za 'mefi. 2. i'mu n'a rā tuya rā zi mane. 3. i geetwū rā besinu ko rā 'mehñā nu'ā rā zi ñohō. 4. nde handāthō nu'ā rā zi tiya eenbi ko nu'ā rā 'mehñā'ā rā zi ñ'ohō di pa rā 'memapa. 5. i maa rā zi 'mefi'ā rā zi ñ'ohō mi šudi. 6. bi m'efa'ā rā zi 'mehñā bi hñātsā rā ñhuni. 7. di ge'ā bōs yaa maa bi 'wini'ā rā dāme. 8. 'ne bi pengi pa 'mū rā ngu nu'ā rā 'mehñā pa 'wingyā mbo'oni nu rā nguu.

Literal English Translation

1. a little man the little laborer has gone to his little work. 2. and there is a the old lady his little *comadre*. 3. and nearby his neighbor with his wife this the little man. 4. well, she sees only this the little old lady that is to say with this his wife this the little man he goes to the laborer. 5. and he went to the little work the little man in the morning. 6. she went after the little woman she brought his meal. 7. from this well already she goes to give to eat to her husband. 8. then she goes back to stay in her house this the woman to give to eat to her animals this her house.

Free English Translation

1. Once there was a nice little *peon* who went to work. 2. His *comadre* lived near this man and his wife, and she could see what the wife was doing. The man would go to work and his wife would follow him later and bring his meal. 8. Then she would go back to her home and feed her animals.

Figure 17.1. Excerpt from Ñähñu stories.

SOURCE: H. R. Bernard and J. Salinas Pedrazas, "Otomí Parables, Folktales, and Jokes," *International Journal of American Linguistics*, Native American Text Series, Vol. 1, no. 2, © 1976 by The University of Chicago Press. All rights reserved. Reprinted with permission.

ogists try to understand and present to an audience the meaning of cultural utterances and practices. Tedlock's work with Andrés Xiloj, Sherzer's with Alberto Campos, and my work with Jesús Salinas are experiments in method to understand the grammar of narrative performance.

Sherzer and Woodbury (1987) observed that highly artistic, creative performance may be based on an underlying cognitive representation—a schema for performance that goes beyond the lexical and syntactic levels of grammar—and you can discover schemas by systematically comparing texts across performances. It is to schemas that we turn next.

Schemas, Models, and Metaphors

Schema analysis combines elements of anthropological linguistics and cognitive psychology in the examination of text. For example, we hear sentences

every day that we've never heard before and somehow we manage to decode them. We don't have a list of sentences in our heads. Instead, we learn a list of rules for making words and for putting words together into sentences.

Some rules are phonological. Consider this sentence: "He worked for two bosses at the same time." We don't pronounce the word "bosses" as if it were "bossiss" (rhymes with the second syllable in "practice"). That would violate the phonological rule that demands voicing of sibilants (like the final *s* in *bosses*) after vowels like the *ɪ* (read: barred-*i* or barred-*eye*) in "bosses." When you add voice to the *s* sound, it becomes a *z* sound.

Some rules are syntactic. We don't say "He is writing book" because that violates the English syntactic rule that requires an article (either "the" or "a") before the noun "book."

And some rules are semantic. We don't say "busy, purple forests dream indignantly" because, even though the syntax is correct, that would violate semantic rules about the kinds of things that can be busy or purple or that can dream. It is, however, the prerogative—even the mandate—of poets to concoct new images by violating just these rules.

Phonology, syntax, and semantics are increasingly complex sets of rules for building sensible utterances. Schema analysis takes this rule-based explanation of behavior a step further. Everyday life—to say nothing of special situations, like major rituals—is just too complex for people to deal with one scene at a time. There must be, the reasoning goes, some rules—a grammar—that help us make sense of so much information. These rules comprise schemas (Casson 1983:430).

Schemas, or **scripts**, as Schank and Abelson (1977) called them, enable culturally skilled people to fill in the details of a story. We often hear things like "Fred lost his data because he forgot to save his work." We know that Fred's forgetting to save his work didn't actually *cause* him to lose his data. A whole set of links are left out, but they are easily filled in by listeners who have the background to do so.

When you buy a car, you expect to bargain on the price, but when you order food in a restaurant you expect to pay the price on the menu. You know that you are supposed to tip in certain kinds of restaurants and that you don't tip at fast-food counters. When someone you hardly know says, "Hi, how's it going?" they don't expect you to stop and give them a complete run-down on how your life is going these days. If you did launch into a peroration about your life, you'd be acting outside the prevailing schema—breaking frame, as Erving Goffman put it (1974). When people do that, we react viscerally and wonder "How the heck did they get in here with all the sane people?"

When many people in a society share a schema, then the schema is cultural. How can we learn about cultural schemas? Most anthropologists do this by

analyzing narratives. Willett Kempton (1987), for example, asked people to tell him about how they adjusted the thermostats for the furnaces in their homes. He found that Americans have two quite different schemas for how thermostats work. Some people hold to a feedback theory: the thermostat senses the temperature and turns the furnace on or off in order to keep the room at some desired temperature. This theory produces set-it-and-forget-it behavior. You set the thermostat at some temperature and let the system do its job. Other people hold to a valve theory. You set the thermostat at some much higher temperature than what you really want. This forces the furnace to pour out lots of heat, fast. When the temperature is where you want it, you turn the dial down. The first theory is etically correct, but the second is widely held and is responsible for a lot of wasted energy. (People who push the elevator button over and over again probably subscribe to a valve theory. We could test that.)

Three Examples of Schema Analysis

The Trobriand land dispute schema. Edwin Hutchins (1980) recorded and transcribed a formal dispute in the Trobriand Islands, in which two men, Motobasi and Kailimila, make speeches to an open court in their village. In these speeches, both men claim the right to cultivate a particular garden plot.

Motobasi, it seems, has sent one of his followers to cut a garden. Kailimila disputes Motobasi's right to cut the garden. Motobasi says:

It is easy for me to take up this garden and cut it. I was cutting it when my younger brothers said "you have recently come. You shall not touch these things. These are our things because we *pokala*'ed previously." But as you know, this was a woman's garden, Ilawokuvamalagi's garden. My older brother cut it by himself. When he died, he gave it to his sister. (ibid.:68)

To win the case, Motobasi will have to refute the statements of his younger brothers who tell him "You shall not touch these things." To understand Motobasi's logic, we need to know that *pokala* is the giving of something by someone of inferior status to someone of superior status "in the hope, but without the promise, that something will be returned" (ibid.:25–26) and we need to know that Trobriand society is divided into matrilineal descent groups.

Motobasi's claim on the garden depends on his listeners filling in around the edges. He hopes that his matrilineal claim, through his sister, will trump his younger brothers' claim through *pokala*. Eventually, the fact that Motobasi could not specify the person whom the brothers had *pokala*'ed will prejudice his case.

The most important finding in Hutchins's account, though, is not the fact that the Trobrianders have a schema for dealing with property disputes and that their schema—with its matrilineal clans and *pokala* exchange, and so on—is very different from our own. Hutchins shows that, within their schema, the Trobrianders use the same logic as we would use in coming to a decision in a dispute. This implies that the rules of logic are universal and it lays to waste the idea that technologically primitive peoples are not up to the abstract logic that Westerners are so proud of inventing (ibid.:128).

The American marriage schema. Naomi Quinn interviewed 11 American couples about marriage. The couples came from different parts of the country. Some were recently married, others were married a long time. And they represented various occupations, education levels, and ethnic and religious groups. Each of the 22 people were interviewed separately for 15 to 16 hours, and the interviews were transcribed.

Quinn has analyzed this body of text to discover the concepts underlying American marriage and to show how these concepts are tied together—how they form a cultural schema, shared by people from different backgrounds about what constitutes success and failure in marriage (Quinn 1982, 1987, 1992, 1996, 1997).

Quinn's method is to look for metaphors in rhetoric and to deduce the schemas, or underlying principles, that could produce those metaphors. For instance, Quinn's informants often compared marriages (their own and those of others) to manufactured and durable products ("It was put together pretty good") and to journeys ("We made it up as we went along; it was a sort of do-it-yourself project"). And when people were surprised at the breakup of a marriage, they would say things like "That marriage was like the Rock of Gibraltar" or "It was nailed in cement." People use these metaphors because they assume that their listeners know that cement and the Rock of Gibraltar are things that last forever.

The method of looking at metaphors as indicators of schemas was developed by George Lakoff and Mark Johnson (2003 [1980]), but Quinn goes further. She reasons that if schemas are what make it possible for people to fill in around the bare bones of a metaphor, then the metaphors must be surface phenomena and cannot themselves be the basis for shared understanding. She tries to understand how metaphors group together and finds that the hundreds of metaphors in her enormous corpus of text all fit into just eight classes: last-ingness, sharedness, compatibility, mutual benefit, difficulty, effort, success (or failure), and risk of failure.

The classes of metaphors, the underlying concepts, are linked together in a schema that guides the discourse of ordinary Americans about marriage. Here is Quinn's understanding of that schema:

Marriages are ideally lasting, shared and mutually beneficial. . . . Benefit is a matter of fulfillment. . . . Fulfillment and, more specifically, the compatibility it requires, are difficult to realize but this difficulty can be overcome, and compatibility and fulfillment achieved, with effort. Lasting marriages in which difficulty has been overcome by effort are regarded as successful ones. Incompatibility, lack of benefit, and the resulting marital difficulty, if not overcome, put a marriage at risk of failure. (Quinn 1997:164)

The Mexican folktale schema. Holly Mathews (1992) collected 60 tellings of *La Llorona* (the weeping woman), a morality tale told across Mexico. Here is one telling, which Mathews says is typical:

La Llorona was a bad woman who married a good man. They had children and all was well. Then one day she went crazy and began to walk the streets. Everyone knew but her husband. When he found out he beat her. She had much shame. The next day she walked into the river and drowned herself. And now she knows no rest and must forever wander the streets wailing in the night. And that is why women must never leave their families to walk the streets looking for men. If they are not careful they will end up like La Llorona. (p. 128)

In another telling, La Llorona kills herself because her husband becomes a drunk and loses all their money. In yet another, she kills herself because her husband is seen going with other women and La Llorona, in disbelief, finally catches him paying off a woman in the streets.

Mathews found that men and women tended to emphasize different things in the story, but the woman always winds up killing herself, no matter who tells it. The morality tale succeeds in shaping people's behavior, she says, because the motives of the characters in the story conform to a schema, shared by men and women alike, about how men and women see each other's fundamental nature (ibid.:129).

Men, according to Mathews's understanding of the cultural model in rural Mexico, view women as sexually uncontrolled. Unless they are controlled, or control themselves, their true nature will emerge and they will begin (as the story says) to "walk the streets" in search of sexual gratification. Men, for their part, are viewed by women as sexually insatiable. Men are driven, like animals, to satisfy their desires, even at the expense of family obligations. In her grammar of the La Llorona tales, Mathews shows that women have no recourse but to kill themselves when they cannot make their marriages work.

Mathews goes beyond identifying the schema and tries to explain where the schema comes from. Most marriages in the village where Mathews did her research (in the state of Oaxaca) are arranged by parents and involve some exchange of resources between the families. Once resources like land are

exchanged there's no turning back, which means that parents can't, or won't take back a daughter if she wants out of a marriage. Then, as Mathews explains, the only way a woman can end her marriage is suicide (ibid.:150). And that, Mathews, says, is why suicide is part of virtually all tellings of the La Llorona tale. (There is an enormous literature on schema analysis. For the history of this key concept in psychology, see Brewer [1999, 2000] and Nishida [1999]. For more on the application of the schema method in anthropology, see Holland and Quinn 1987, D'Andrade 1991, D'Andrade and Strauss 1992, D'Andrade 1995:chap. 6, and Strauss and Quinn 1997.)

Discourse Analysis

Discourse analysis involves the close study of naturally occurring interactions. Scholars in this arena, like Brenda Farnell and Laura Graham (1998:412), see discursive practices—all the little things that make our utterances uniquely our own—as concrete manifestations both of the culture we share with others and of our very selves. What we do in traditional ethnography, according to Joel Sherzer and Greg Urban (1986:1), is distill discourse into a description of a culture. What we ought to be observing—and describing—they say, is ordinary discourse itself, since culture emerges from the constant interaction and negotiation between people.

Here's an example of how much of culture can be encapsulated in an ordinary discourse event. I went to Japan for 4 months in 1991 to work at the National Museum of Ethnology in Osaka. I didn't speak Japanese, so I was limited to studying things that I could observe directly and about which I could hold discussions, in English, with my colleagues at the museum. I noticed that many Japanese bowed when they said good-bye to someone on the telephone and I became interested in the cultural artifact of bowing.

I rode the bus to work every day. Most regular riders of the buses in Osaka buy blocks of tickets in advance, at a discount. When people leave the bus, they drop one of their fare tickets into a hopper next to the driver. As the bus rolled to a stop one day, a boy of about 5 implored his mother to let him drop off her ticket as they left the bus. The mother gave him the ticket, left the bus, and watched from the sidewalk as the toddler dropped off the ticket and scampered, grinning, down the bus's stairs to her.

The mother gave the child a stern look and said something that I couldn't hear. The child scampered back up the stairs, faced the driver, and bowed deeply from the waist. "*Arigato gozaimashitá,*" said the boy to the driver, with heavy emphasis on the final syllable. The driver bowed slightly and

acknowledged the boy, saying “*Doozu*,” while smiling at the mother who waited at the bottom of the stairs for her son.

It turns out that the child was mimicking, in phrase and diction, the formal way that an adult man might say thank you with great emphasis in Japanese. Had I been able to tape that one small discourse event, I’m sure that there would have been enough exegesis by my Japanese colleagues at the museum to have produced a book on socialization of men and women into the ways of bowing and of saying thanks.

Formal discourse analysis, in fact, involves taping interactions and careful coding and interpretation. Howard Waitzkin and his colleagues (1994:32) taped 336 encounters involving older patients and primary-care internists. (The physicians included some in private practice, some who worked in a teaching hospital, and some who worked both in private practice and in hospital outpatient departments.)

Waitzkin et al. randomly chose 50 of these encounters for intensive study. They had each of the 50 encounters transcribed, verbatim—with all the “uhs,” pauses, and whatnot that occur in real discourse—and two research assistants checked the accuracy of the transcriptions against the original tape.

During the interpretation phase, research assistants read the transcripts and noted the ones that dealt with aging, work, gender roles, family life, leisure, substance use, and socioemotional problems—all areas in which the researchers were interested at the time. The assistants read through the transcripts and tagged instances where either the doctors or the patients “made statements that conveyed ideological content or expressed messages of social control” (ibid.:328).

To illustrate their interpretive approach to discourse analysis, Waitzkin et al. go through two texts in detail. They use the same method as that used in biblical exegesis: A chunk of text is laid out, followed by commentary involving all the wisdom and understanding that the commentators can bring to the effort. For example, Waitzkin et al. produce this snippet of interaction between a doctor (D) and his patient (P), an elderly woman who has come in for a follow-up of her heart disease:

P: Well I should—now I’ve got birthday cards to buy. I’ve got seven or eight birthdays this week—month. Instead of that I’m just gonna write ’em and wish them a happy birthday. Just a little note, my grandchildren.

D: Mm hmm.

P: But I’m not gonna bother. I just can’t do it all, Dr.—

D: Well.

P: I called my daughters, her birthday was just, today’s the third.

D: Yeah.

P: My daughter's birthday in Princeton was the uh first, and I called her up and talked with her. I don't know what time it'll cost me, but then, my telephone is my only connection.

Waitzkin et al. comment:

At no other time in the encounter does the patient refer to her own family, nor does the doctor ask. The patient does her best to maintain contact, even though she does not mention anything that she receives in the way of day-to-day support. Compounding these problems of social support and incipient isolation, the patient recently has moved from a home that she occupied for 59 years. (ibid.:330–331)

When they get through presenting their running commentary on the encounter, Waitzkin et al. interpret the discourse:

This encounter shows structural elements that appear beneath the surface details of patient-doctor communication. . . . Contextual issues affecting the patient include social isolation; loss of home, possessions, family, and community; limited resources to preserve independent function; financial insecurity; and physical deterioration associated with the process of dying. . . . After the medical encounter, the patient returns to the same contextual problems that trouble her, consenting to social conditions that confront the elderly in this society.

That such structural features should characterize an encounter like this one becomes rather disconcerting, since the communication otherwise seems so admirable. . . . The doctor manifests patience and compassion as he encourages a wide-ranging discussion of socioemotional concerns that extend far beyond the technical details of the patient's physical disorders. Yet the discourse does nothing to improve the most troubling features of the patient's situation. To expect differently would require redefining much of what medicine aims to do. (ibid.:335–336)

Waitzkin et al. make clear that alternative readings of the same passage are possible and advocate, as part of their method, the systematic archiving, in publicly available places, of texts on which analysis is conducted. When the project was over, Waitzkin filed the transcripts with University Microfilms International so that other researchers could use the data for later analysis.

Conversation Analysis

If you listen carefully to ordinary conversations between equals, you'll hear a lot of sentence fragments, false starts, interruptions, overlaps (simultaneous speech), and repeating of words and phrases. It may sound messy at first, but

as students of conversation have learned, there is order in all that seeming chaos as participants respond to each other (even in strong disagreements and shouting matches) and take turns (Goodwin 1981:55ff). **Conversation analysis** is the search for the grammar of ordinary discourse, or **talk-in-interaction**. It is the study of how people take turns in ordinary discourse—who talks first (and next, and next), who interrupts, who waits for a turn.

The rules of turn-taking are, like the rules of grammar that govern the formation of sentences, known to native speakers of any language. But unlike the other rules of grammar, the rules for taking turns are flexible, and allow turn-taking to be negotiated, on the fly, by participants in a conversation. At the molecular level, then, every conversation is unique, but the study of many conversational exchanges can expose the general rules, within and across cultures, that govern how conversations start, evolve, and end.

TRANSCRIPTIONS

In order to identify turns and other features of conversations, you need detailed records of actual talk-in-interaction. The tactic for signaling the intention to take a turn or to repair a broken turn sequence may involve words or it may involve prosodic features of speech (intonation, length of vowels, stress, and so on), or even involve breaths, tokens (like er, ummm, eh), or gestures or gazes. Coding schemes have been devised for marking all these different features.

The system most widely used for transcribing speech was developed by Gail Jefferson (1973; see Sacks et al. 1974; Psathas 1979; Atkinson and Heritage 1984; Psathas and Anderson 1990). Conventions for recording gestures and gazes were developed by Charles Goodwin (1994). Table 17.1 provides the common conventions for transcribing conversations.

TAKING TURNS

Harvey Sacks and his colleagues, Emmanuel Schegloff and Gail Jefferson, are widely credited for developing the systematic study of order in conversations (Schegloff 1972; Jefferson 1973; Schegloff and Sacks 1973; Sacks et al. 1974). Among the basic rules they discovered are that the person who is speaking may (but does not have to) identify the next speaker to take a turn (Sacks et al. 1974). This is done with conversational devices—like “so what do you think, Jack?”—or with gazes or body language (Goodwin 1986, 1994). If the person speaking does not select the next speaker, then any other person in the conversation can self-select to take a turn.

Alternatively, the next speaker may jump in before a speaker has completed

TABLE 17.1
Symbols for Transcribing Verbatim Text

<i>Symbol</i>	<i>Definition</i>
:	Indicates length. More colons indicate more length.
>text<	Speech between angle brackets is faster than normal speech.
(text)	Text in parentheses means that the transcriber had doubts about it.
()	Parens with no text means that transcriber could not make it out at all.
(1.6)	Numbers in parens indicate pauses, in seconds and tenths of a second
(.)	A period in parens indicates an untimed and quick pause
((text))	Double parens contain comments by the researcher about people's gestures, gazes, and so on. Often in italics.
[]	Square brackets indicate where one person interrupts another or talks simultaneously.
–	An en-dash (longer than a hyphen) indicates an abrupt end in the middle of a word
?	A question mark indicates rising intonations. It does not necessarily mean that a question is being asked.
.	A period indicates falling intonation. It does not necessarily mean the end of a sentence.
=	The equal sign indicates that a person takes a turn immediately as the previous turn ends.
°text°	Text between degree symbols is quieter than the rest of the text.
TEXT	Text in caps is louder than the rest of the text.
<u>Text</u>	Underlining means that the text is emphasized.
.hh and hh	These indicate inhaled (with period first) and exhaled (no period) breaths, as often occurs in natural conversation.

SOURCE: John F. Manzo, "Taking Turns and Taking Sides: Opening Scenes from Two Jury Deliberations," *Social Psychology Quarterly* 59, p. 124 (June 1996). Reproduced by permission of the American Sociological Association.

a turn, by anticipating the end of a turn—a supportive gesture—or by interrupting and trying to take away the speaker's turn by force—a hostile gesture. Big gaps occur so rarely in real conversations because speakers anticipate the end of turns so well. For example:

1. A: what a cute *nose*.
2. B: all babies ha–
3. A: [no they don't

A takes a turn; B jumps in immediately; and then A says, "no they don't" just as the syllable "ba" in "baby" registers. There are no gaps because the end of the second turn, by B, is predictable to A.

If the turn-taking runs out of steam in a conversation, either because it

becomes unpredictable or the content gets used up, and no one jumps in to take a turn at the appropriate time, then the current speaker may (but does not have to) continue talking. If none of these things happen, then there will be a gap in the conversation. Gaps don't usually last very long if the rules of conversation are followed.

Of course, the rules are often broken in real conversations. People often do interrupt each other unsupportively and don't wait to take their turns. People who get interrupted don't always push on, trying to finish their turn, but relinquish their turn instead, without finishing. People usually recognize when the preferred order of turn-taking has not been adhered to and engage in what are called repair tactics. For example, Sacks et al. (1974) noted that when two people start talking over each other, one of them might just stop and let the other finish.

1. A: [It's not like we–
2. B: [Some people are–
3. B: Sorry.
4. A: No, g'head
5. B: hhhh I wus jus gonna say that some people aren't innerested in sports at all.
6. A: [Right

This is a simple repair tactic that doesn't involve any serious content. Other repair tactics can be more complex. The result of this repair sequence could have come out differently, of course. But in the study of hundreds of natural conversations between equals, we can uncover the rules governing how people open and close conversations, how they repair mistakes (when they break the rules of turn-taking, for example), and how they segue from one theme to another.

The turn-taking sequence can be suspended, as for example, in the telling of jokes and stories. If I say, "Did you hear the one about? . . ." and you say, "No, tell me," then this suspends turn-taking until you're finished with the joke. If you interrupt me in the middle of telling the joke, this breaks the rule for this conversation element. And the same thing goes for storytelling. If I say, "I was on the flight from Hell coming back from Detroit last week" and you respond by saying "What happened?" then I get to tell you the story all the way through. Certain kinds of interruptions are permitted, but sidetracking me completely from the story is not permitted.

Here, again, though, we find the rule broken all the time. And when sidetracking happens, one of several repair sequences might kick in. "Sorry, I got you off track. Then what happened?" is a repair sequence. We've all experi-

enced the pain of never getting back to a story from which we were sidetracked in a conversation. When that happens, we might think badly of the person who did it, or we might shrug it off—depending on the context and what’s at stake. If you’re in a job interview and the interviewer sidetracks you, you’d probably think twice about insisting that the interviewer let you finish the story you were telling.

ADJACENCY PAIRS

Among the first things that conversation analysts noticed when they started looking carefully at conversations was ordered pairs of expressions, like questions and greetings (Schegloff and Sacks 1973). Once the first part of a pair occurs, the second part is expected.

Sacks (1992:3ff) noticed that workers at a psychiatric hospital’s emergency telephone line greeted callers by saying something like, “Hello. This is Mr. Smith. May I help you?” Most of the time, the response was “Hello, this is Mr. Brown,” but on one occasion, the caller responded, “I can’t hear you.” When the worker repeated his greeting, “This is Mr. *Smith*,” with an emphasis on Smith, the caller responded “Smith.” In this case, the rule for an adjacency pair was not really being broken. It was being negotiated by both parties, on the fly, during the conversation. Mr. Smith, the suicide prevention worker, was trying to get the caller to give his name, and the caller was trying *not* to give his name.

If conversations are dynamic things, with constant negotiation, then, to understand the rules of conversations, you can’t interview people about the rules of, say, greetings. You have to study real conversations in which greetings occur.

Here is an example of an adjacency pair—a question-answer pair—that’s broken up between its parts.

1. A: Is it good? The Szechuan pork?
2. B: Y’like spicy, uh—
3. A: [Yeah
4. B: It’s kinda, y’know, hot.
5. A: *Great*
6. B: Yeah, me too.

There is a lot going on here. A asks B if a particular dish is good. This is the first part of a question pair, which calls for an answer. B responds with another question, which creates a new expectation for an answer. A anticipates this and answers “yeah” (line 3) before B finishes her thought (which presumably was “y’like spicy, uh, food?”).

B doesn't want to give up her turn just yet, so she completes the thought, in line 4. This gives A the opportunity to reaffirm that she does, in fact, like spicy food (that her "yeah" in line 3 was not just an acknowledgment of B's question) before she answers the first one. A reaffirms her "yeah" response with "great" in line 5. The segment ends with B agreeing that she, too, likes spicy food.

Looking at this segment, it seems like the first adjacency pair, the question about whether the Szechuan pork is good, has never been answered. But it has. The answer is implied in the inserted question in line 2 and the retesting in line 4, all of which is a shorthand for: "Yes, if you like spicy food, then the Szechuan pork is good; otherwise it isn't."

If this seems complicated, it's because it is. It's very, very complicated. Yet, every native speaker of English understands every piece of this analysis because: (1) The segment of conversation is rule based; (2) We understand the rules; and (3) We understand that the outcome is the result of cooperation by both speakers, A and B, as they attend to each other's words, draw on cultural knowledge that they share, and build a conversation dynamically. In this case, two people, in a conversation among equals, have worked together to build a particular outcome.

And if you think this is complicated, just add more speakers. With two people in a conversation, A and B, there are two pairs of people—AB (as seen from A's perspective) and BA (as seen from B's perspective)—figuring out each other's motives and behaviors and negotiating their way through the interaction. With three people in a conversation (A, B, and C), there are six pairs (AB, BA, AC, CA, BC, CB) doing the negotiating. With six people, there are 30 such pairs. Deborah Tannen (1984) analyzed 2 hours and 40 minutes of conversation among six friends at a Thanksgiving dinner. The analysis alone is a 170-page book, and that doesn't include the hundreds of pages of transcripts.

Dinner parties often start out as conversations among six or eight simultaneous participants, but they often break up into smaller conversations that are easier to manage. Some conversations among large groups of people, however, require special rules so that they don't break up, like dinner parties do, into free-for-alls. Juries, for example, have 6, 9, 12, or 21 members, depending on the kind of trial, and are a special kind of conversation. They have a leader (the jury foreperson), and people work hard to make sure that the cultural rules for such formal conversations (written down nowhere, but widely understood) are followed. Among other things, the rules require that no one gets completely drowned out and sidelined, unless they deserve to be (by first breaking the cultural rules themselves and trying to dominate). (See Manzo [1996] for a detailed analysis of turn-taking by a jury.)

We come next to the two methods that are very widely used across the social sciences for analyzing text: grounded theory and classical content analysis. Grounded theory emphasizes the discovery and labeling of concepts (variables) and the building of models based on a close reading of the text. Content analysis emphasizes the formal description of concepts and the testing of models and hypotheses.

Grounded Theory

The **grounded-theory** approach is a set of techniques for: (1) identifying categories and concepts that emerge from text; and (2) linking the concepts into substantive and formal theories. The approach was developed by sociologists (Glaser and Strauss 1967; Strauss and Corbin 1990) and is widely used to analyze ethnographic interview data.

The mechanics of grounded theory are deceptively simple: (1) Produce transcripts of interviews and read through a small sample of text. (2) Identify potential **analytic categories**—that is, potential **themes**—that arise. (3) As the categories emerge, pull all the data from those categories together and compare them. (4) Think about how categories are linked together. (5) Use the relations among categories to build theoretical models, constantly checking the models against the data—particularly against negative cases. (6) Present the results of the analysis using **exemplars**, that is, quotes from interviews that illuminate the theory.

The key to making all this work is called **memoing**. Throughout the grounded-theory process, you keep running notes about the coding and about potential hypotheses and new directions for the research. Grounded theory is an **iterative process** by which you, the analyst, become more and more *grounded* in the data. During the process, you come to understand more and more deeply how whatever you're studying really works.

Coding Themes

The heart of grounded theory is identifying themes in texts and coding the texts for the presence or absence of those themes. Coding turns free-flowing texts into a set of nominal variables (what statisticians mean, by the way, when they use the phrase “qualitative data”). In a set of texts about the experience of divorce, people do or do not talk about what to do about pension funds; they do or do not talk about how their children are taking it; they do or do not talk about their relations with their former in-laws; and so on.

Where do you stop? There is practically no end to the number of themes

you can isolate for any text. When I was in high school, my physics teacher put a bottle of Coca-Cola on his desk and challenged our class to come up with interesting ways to describe that bottle. Each day for weeks that bottle sat on his desk as new physics lessons were reeled off, and each day new suggestions for describing that bottle were dropped on the desk on the way out of class.

I don't remember how many descriptors we came up with, but there were dozens. Some were pretty lame (pour the contents into a beaker and see if the boiling point was higher or lower than that of sea water) and some were pretty imaginative (cut off the bottom and test its magnifying power), but the point was to show us that there was no end to the number of things we could describe (measure) about that Coke bottle, and the point sunk in. I remember it every time I try to code a text.

Inductive and Deductive Coding

Grounded-theory research is mostly based on **inductive or “open” coding**. The idea is to become grounded in the data and to allow understanding to emerge from close study of the texts. That's why Barney Glaser and Anselm Strauss (1967), the originators of this approach, titled their book *The Discovery of Grounded Theory*.

Content analysis (which we'll take up in the next section) is mostly based on **deductive coding**. In doing deductive analysis of text, you start with a hypothesis *before* you start coding. The idea is to test whether your hypothesis is correct. Inductive research is what you do when you're in the **exploratory and discovery phase** of *any* research project, whether your data are words or numbers. Deductive research is what you do in the **confirmatory** stage of any research project—no matter what kind of data you have. There is no point in talking about which is better. They're both terrific if you use them to answer appropriate questions.

How to Find Themes

For inductive research, like grounded theory, you can't go wrong by just reading the texts and underlining or highlighting things as you go. You can get fancy later (more about this in a minute), but when you start to work with a corpus of written text, just read it and if you see something that you think might be important, highlight it. Some of the words and phrases you highlight will turn into names for themes. In fact, Strauss and Corbin (1990:68) recommend explicitly using actual phrases your text—the words of real people—to name themes, a technique they call **in vivo** coding.

Willms et al. (1990) and Miles and Huberman (1994) suggest starting with some general themes derived from reading the literature and adding more themes and subthemes as you go. This is somewhere between inductive and deductive coding. You have a general idea of what you're after and you know what at least some of the big themes are, but you're still in a discovery mode, so you let new themes emerge from the texts as you go along.

Look for repetitions. "Anyone who has listened to long stretches of talk," says Roy D'Andrade, knows how frequently people circle through the same network of ideas" (1991:287). In my study of how ocean scientists interact with the people in Washington, D.C., who are responsible for ocean policy (Bernard 1974), I kept hearing the word "brokers." Scientists and policymakers alike used this word to describe people whom they trusted to act as go-betweens, so "broker" became one of the code themes for my work.

Look for unusual terms or common words that are used in unusual ways. James Spradley (1972) recorded conversations among homeless men (they were called tramps in those days) at informal gatherings, meals, and card games. Spradley kept hearing the men talk about "making a flop," which was their jargon for finding a place to sleep each night. Spradley went through his material and isolated everything he could find about flops: ways to make a flop, kinds of people who bother you when you flop, and so on. Then Spradley went back to his informants and asked them for more information about each of these subthemes.

And, said Spradley (1979:199–201), look for evidence of social conflict, cultural contradictions, informal methods of social control, things that people do in managing impersonal social relationships, methods by which people acquire and maintain achieved and ascribed status, and information about how people solve problems. Each of these arenas is likely to yield major themes in cultures.

More Techniques for Finding Themes

Other techniques for finding themes include pile sorting, making word counts, and producing key-word-in-context, or KWIC, tables. To use the **pile-sorting** method—what Lincoln and Guba (1985:347ff) call cutting and sorting—look for real quotes from the interviews you do with informants that represent what you think are important topics in the data. Cut out each quote (making sure to leave some of the context in which the quote occurs) and paste it onto a 3" × 5" index card. On the back of the card, note who said it and where it appeared in the text. Then, lay out the cards on a big table, sort them into piles of similar quotes, and *name each pile*. These are the themes (Ryan and Bernard 2003:94).

In his research on household healthcare in Njinikom, Cameroon, Gery Ryan (1995) had several coders sorting the quotes-on-cards into piles. He recorded their conversation as they discussed what they were doing and used the discussion to help identify the themes in his data. Later, in chapter 21, I'll introduce you to multidimensional scaling (MDS). We use MDS to find similarities in pile sorts, and there is no reason you can't do that with your pile sort of your own data.

Word lists and the **KWIC technique** are extensions of the philosophy behind *in vivo* coding in grounded theory: If you want to understand what people are talking about, look closely at the words they use. The method has a very, very long history. The classic KWIC method is a **concordance**, which is a list of every substantive word in a text with its associated sentence. Concordances have been done on sacred texts from many religions and on famous works of literature from Euripides (Allen and Italic 1954), to Beowulf (Bessinger and Smith 1969), to Dylan Thomas (Farrington and Farrington 1980). (See McKinnon [1993] and Burton [1981] on the use of concordances in modern literary studies.)

It's easier to build concordances today than it was before computers, and many programs are available that will list all the unique words in a text and count how many times each is used (see appendix F).

Ryan and Weisner (1996) asked fathers and mothers of adolescents, "Describe your children. In your own words, just tell us about them." In looking for themes in these rich texts, Ryan and Weisner did a word count. Mothers were more likely than fathers to use words like "friends," "creative," and "honest"; fathers were more likely than mothers to use words like "school," "student," and "independent." These counts became clues about the themes that Ryan and Weisner eventually used in coding the texts.

There is another technique, called semantic network analysis, which involves tracking the association among words or concepts in a text. This method, though, like pile sorting, also requires multidimensional scaling, so I'll hold off on telling you about it until we get to chapter 21.

No matter how you actually *do* inductive coding—whether you start with paper and highlighters or use a computer to paw through your texts; whether you use *in vivo* codes, or use numbers, or make up little mnemonics of your own; whether you have some big themes in mind to start or let all the themes emerge from your reading—by the time you identify the themes and refine them to the point where they can be applied to an entire corpus of texts, a lot of interpretive analysis has already been done. Miles and Huberman say simply: "Coding is analysis" (1994:56).

Kurasaki's and Nyamongo's Codebooks

Kurasaki (1997) studied the ethnic identity of *sansei*, third-generation Japanese Americans. She interviewed 20 people and used a grounded-theory approach to do her analysis. She started with seven major themes: (1) a sense of history and roots, (2) values and ways of doing things, (3) biculturality, (4) sense of belonging, (5) sense of alienation, (6) self-concept, and (7) worldview. As the analysis progressed, she split the major themes into subthemes. So, for example, she split the first theme into (1) sense of having a Japanese heritage and (2) sense of having a Japanese American social history.

As the coding progressed further, she eventually decided to combine two of the major themes (sense of belonging and sense of alienation) and wound up with 6 major themes and a total of 18 themes. Kurasaki assigned her own numerical codes to each of the themes—1.1 for the first subtheme in macrotheme 1, 7.2 for the second subtheme in macrotheme 7, and so on—and used those numbers to actually code her transcribed interviews. Her coding scheme is shown in table 17.2.

Nyamongo (1998) did semistructured interviews with 35 Gusii people in Kenya about how they responded to various symptoms associated with malaria. In addition to using a grounded-theory approach to develop the theme codes, Nyamongo wanted to do statistical analysis of his data. Table 17.3 shows his codebook. (Nyamongo did all the interviewing in Gusii—a language spoken by about two million people in Kenya—and all the interviews were transcribed in Gusii, but he coded the transcripts using English themes and English-looking mnemonics.)

Notice the difference between Nyamongo's and Kurasaki's codebooks. In addition to some basic information about each respondent (in columns 1–33), Nyamongo coded for 24 themes. Those 24 themes, however, were nominal variables. He coded occupation, for example, as one of eight possible types. And notice that he used words, not numbers for those occupations, and that one possible entry was “n.a.” This is a common abbreviation used in research for “information not available.”

The code SYM1, in column 75, stands for “symptom #1” and refers to whether or not an informant mentioned headache in his or her narrative as a symptom of malaria. Nyamongo coded for whether people mentioned any of 14 symptoms, and he added a 15th code, SYM15 (in column 81) for “other.” When you break up a list of things into a series of yes/no, present/absent variables, like Nyamongo did with CAUSE and SYM, and so on, this turns nominal variables into a series of **dummy variables**. Thus, Nyamongo's codebook has a total of 88 theme *variables* (including all the dummy variables), even though he coded for just 24 *themes* (like symptoms, treatments, and so on).

TABLE 17.2
Kurasaki's Coding Scheme for Her Study of Ethnic Identity among Sansei in California

<i>First-Order Category</i>	<i>Second-Order Category</i>	<i>Numeric Code</i>
Sense of history and roots	Sense of having a Japanese heritage	1.1
	Sense of having a Japanese American social history	1.2
Values and ways of doing things	Japanese American values and attitudes	2.1
	Practice of Japanese customs	2.2
	Japanese way of doing things	2.3
	Japanese American interpersonal or communication styles	2.4
	Japanese language proficiency	2.5
Biculturality	Integration or bicultural competence	3.1
	Bicultural conflict or confusion	3.2
Sense of belonging	Sense of a global ethnic or racial community	4.1
	Sense of interpersonal connectedness with same ethnicity or race of others	4.2
	Sense of intellectual connectedness with other ethnic or racial minorities	4.3
	Searching for a sense of community	4.4
Sense of alienation	Sense of alienation from ascribed ethnic or racial group	5.1
Self-concept	Sense of comfort with one's ethnic or racial self	6.1
	Searching for a sense of comfort with one's ethnic or racial self	6.2
Worldview	Social consciousness	7.1
	Sense of oppression	7.2

SOURCE: "Ethnic Identity and Its Development among Third-Generation Japanese Americans" by K. S. Kurasaki, 1997. Ph.D. diss., Department of Psychology, DePaul University.

Table 17.4 shows a piece of the **data matrix** produced by Nyamongo's coding of the 35 narratives.

Building Conceptual Models by Memoing

Once you have a *set of themes* coded in a *set of texts*, the next step is to identify *how themes are linked to each other* in a theoretical model (Miles and Huberman 1994:134–137).

TABLE 17.3
Nyamongo's (1998) Codebook: How Gusii Respond to Malaria

<i>Column</i>	<i>Code</i>	<i>Variable Description and Variable Values</i>
1–3	RESP_ID	Informant number, from 001–035
4–11	NAME	Name of the informant
12–13	AGE	Age in years as reported by the informant.
14–18	GENDER	1 = female, 2 = male
19–32	OCC	Occupation of the informant. This is a nominal variable. The possible entries are housewife (1), farmer (2), retired worker (3), security guard (4), teacher (5), student (6), artisan (7), village elder (8), and n.a. if the informant reported no occupation.
33	SES	Socioeconomic status is an ordinal variable measured by presence of the following in the homestead: 1 = grass-thatched house, 2 = iron-sheet-roofed, dirt-floor house, 3 = iron-sheet-roofed, cemented-floor house, 4 = semipermanent house, 5 = permanent house, 9 = information not available
.	.	.
41	CAUSE1	Mosquito is mentioned by informant as the cause of malaria: 1 = yes, 0 = no
42	CAUSE2	Eating sugarcane, ripe bananas, and roast green maize are mentioned by informant as the cause of malaria: 1 = yes, 0 = no
43	CAUSE3	Causes other than CAUSE1 and CAUSE2 are given: 1 = yes, 0 = no
.	.	.
48	CTRL1	What should be done to reduce malaria cases? 1 = keep compound clean, 0 = other
49	CTRL2	What should be done to reduce malaria cases? 1 = take medicine for prophylaxis, 0 = other
50	CTRL3	What should be done to reduce malaria cases? 1 = use net or spray or burn coil, 0 = other
51	CTRL4	What should be done to reduce malaria cases? 1 = nothing can be done, 0 = other
52	DIAG	Diagnosis of illness by the informant: 1 = malaria, 0 = other illness
53	FACTOR1	Does cost influence whether people use home management or hospital-based care? 1 = yes, 0 = no
54	FACTOR2	Does duration of sickness influence whether people use home management or hospital-based care? 1 = yes, 0 = no

TABLE 17.3 (Continued)

<i>Column</i>	<i>Code</i>	<i>Variable Description and Variable Values</i>
55	FACTOR3	Is intensity (severity) a factor influencing whether people use home management or hospital-based care? 1 = yes, 0 = no
56	HELP1	Type of support given to informant by family or friends: 1 = buying medicine, 0 = other
.		
75	SYM1	The informant mentions headache as a symptom: 1 = yes, 0 = no
.		
80	SYM14	The informant mentions child has unusual cries as a symptom: 1 = yes, 0 = no
81	SYM15	Other—if not one of SYM1 through SYM14: 1 = yes, 0 = no
91	TREAT11	Did the informant report using pills bought over the counter as the first treatment resort? 1 = yes, 0 = no
.		
96	TREAT21	Second treatment resort: The values are the same as for TREAT11, column 91
.		
119	WITCH	The informant mentions that witchcraft may be implicated if patient has malaria: 1 = yes, 0 = no
120	YESDIAG	Does the coder think the informant made the right diagnosis based on the informant's stated symptoms? 1 = yes, 0 = no
121	COST	Cost of receiving treatment from private health care provider in Kenyan shillings, from 0001–9999, 9999 = information not available

SOURCE: Data reproduced by permission of the author.

Memoing is a widely used method for recording relations among themes. In memoing, you continually write down your thoughts about what you're reading. These thoughts become information on which to develop theory. Memoing is taking "field notes" on observations about texts.

Strauss and Corbin discuss three kinds of memos: **code notes**, **theory notes**, and **operational notes** (1990:18, 73–74, 109–129, 197–219). Code

TABLE 17.4
Part of the Data Matrix from Nyamongo's Analysis of 35 Gusii Texts

<i>Person</i>	<i>Age</i>	<i>Sex</i>	<i>Occupation</i>	<i>SES</i>	<i>CTRL1</i>	<i>CAUSE1</i>	<i>HELPI</i>	<i>WITCH</i>
1	48	M	retired worker	2	1	1	0	0
2	58	M	farmer	1	0	1	0	1
3	68	M	clan elder	1	1	1	0	1
.
.
.
33	54	F	teacher	2	1	1	0	0
34	57	F	artisan	2	0	0	0	0
35	26	F	housewife	1	1	1	0	1

notes describe the concepts that are being discovered in “the discovery of grounded theory.” In theory notes, you try to summarize your ideas about what’s going on in the text. Operational notes are about practical matters.

Once a model starts to take shape, start shopping for negative cases—ones that don’t fit the pattern. Suppose you comb through a set of narratives from women in the labor market. Some women say that they got upset with their last job and quit. You find that most of the women who did this have husbands who earn a pretty good living. Now you have a take-this-job-and-shove-it category. Is there a case in which a woman says “You know, I’d be outta this crummy job in a minute if I didn’t have two kids at home to take care of”? Don’t wait for that case to come to you. Go looking for it.

Negative cases either disconfirm parts of a model or suggest new connections that need to be made. In either case, negative cases need to be accommodated when you present your results. (Negative case analysis is discussed in detail by Becker et al. 1961:37–45, Strauss and Corbin 1990:108–109, Lincoln and Guba 1985:309–313, Dey 1993:226–233, and Miles and Huberman 1994:271.)

Displaying Concepts and Models

So, how do you actually build those models and what do they really look like? Here’s a step-by-step example.

Margaret Kearney and her colleagues (Kearney et al. 1995) interviewed 60 women who reported using crack cocaine an average of at least once weekly during pregnancy. The semistructured interviews lasted from 1 to 3 hours and covered childhood, relationships, life context, previous pregnancies, and actions during the current pregnancy related to drug use, prenatal care, and self-care. Kearney et al. coded and analyzed the transcripts as they went. As

new topics emerged, investigators asked about the topics in subsequent interviews. In this way, they linked data collection and data analysis in one continuous effort.

Kearney et al. coded the data first for the general topics they used to guide the interviews. Later, they would use these codes to search for and retrieve examples of text related to various interview topics. Next, team members reread each transcript searching for examples of social psychological themes in the women's narratives. Each time they found an example, they asked: "What is this an example of?" The answers suggested substantive categories that were refined with each new transcript.

Kearney et al. (1995) looked at how substantive categories were related. They recorded their ideas about these interactions in the forms of memos and developed a preliminary model. With each subsequent transcript, they looked for negative cases and pieces of data that challenged their emerging model. They adjusted the model to include the full range of variation that emerged in the transcripts.

To begin with, Kearney et al. identified five major categories, which they called: VALUE, HOPE, RISK, HARM REDUCTION, and STIGMA MANAGEMENT. (Capital letters are often used for code names in grounded-theory research, just as in statistical research.) Women valued their pregnancy and the baby-to-be in relation to their own life priorities (VALUE); women expressed varying degrees of hope that their pregnancies would end well and that they could be good mothers (HOPE) and they were aware that cocaine use posed risks to their fetus but they perceived that risk differently (RISK). Women tried in various ways to minimize the risk to the fetus (HARM REDUCTION) and they used various stratagems to reduce social rejection and derision (STIGMA MANAGEMENT).

By the time they had coded 20 interviews, Kearney et al. realized that the categories HARM REDUCTION and STIGMA MANAGEMENT were components of a more fundamental category that they labeled EVADING HARM. After about 30 interviews had been coded, they identified and labeled an overarching psychological process they called SALVAGING SELF that incorporated all five of the major categories. By the time they'd done 40 interviews, Kearney et al. felt they had reached **theoretical saturation**, which means that they were not discovering new categories or relations among categories. Just to make sure, they conducted another 20 interviews and confirmed the saturation.

Figure 17.2 shows the graphic model that Kearney et al. produced to represent their understanding of how the process worked. Kearney et al. described in rich detail each of the major categories that they discovered, but notice how

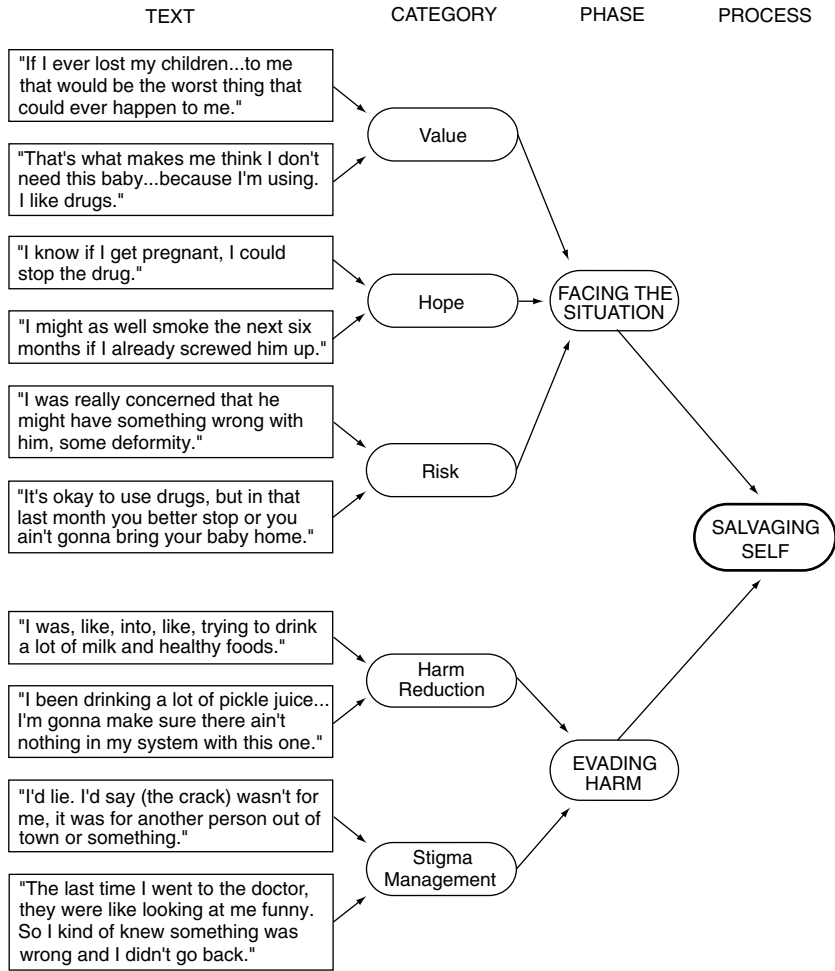


Figure 17.2. The relation of themes/categories in Kearney et al.'s analysis of how pregnant drug users viewed their own behavior.

SOURCE: M. H. Kearney et al., "Salvaging Self—A Grounded Theory of Pregnancy on Crack Cocaine," *Nursing Research*, Vol. 44, pp. 208–13. © 1995, Lippincott Williams & Wilkins. Reprinted with permission.

each of the substantive themes in their model is succinctly defined by a quote from a respondent.

When the steps of the grounded-theory approach are followed, models or theories are produced that are, indeed, *grounded* in the text. These models, however, are not the final product of the grounded-theory approach. In their original formulation, Glaser and Strauss (1967) emphasized that the building of grounded-theory models is a step in the research process. The next, of course, is to confirm the validity of a model by testing it on an independent sample of data. Kearney et al. checked the validity of their model by presenting it to knowledgeable respondents (pregnant drug users), to members of the project staff, and to health and social service professionals who were familiar with the population.

Using Exemplar Quotes

Besides displaying models, one of the most important methods in grounded-theory text analysis is the presentation of direct quotes from respondents—quotes that lead the reader to understand quickly what it took you months or years to figure out. You choose segments of text—verbatim quotes from respondents—as exemplars of concepts and theories or as exemplars of exceptions to your theories (those superimportant negative cases).

This technique looks easy, but it's not. You have to choose the exemplars very carefully because your choices constitute your analysis, as far as the reader is concerned, and you have to avoid what Lofland (1971) called the two great sins of qualitative analysis in order to use the exemplar quote technique effectively.

The first sin, excessive analysis, involves the all-too-familiar practice of jargon writing and the avoidance of plain English to say plain things. If you analyze a batch of data and conclude that something simple is going on, don't be afraid to say so. There is absolutely nothing of scientific value to be gained from making straightforward things complicated.

Compare these two sentences: (1) "The more generations that people from various ethnic groups are in the United States, the less likely they are to speak anything but English." (2) "Over an expanding number of generations, people of ethnic heritage in the United States become, probabilistically, less likely to adhere to their traditional linguistic symbol systems." The best word to describe the second sentence is *yucky*.

The second sin consists of avoiding doing any analysis on your own—being so gun-shy of theory and jargon that you simply fill up your papers and books with lengthy quotes from people and offer no analysis at all. Data do not speak for themselves. You have to develop your ideas (your analysis) about what's

going on, state those ideas clearly, and illustrate them with selected quotes from your respondents.

Katherine Newman (1986), for example, collected life history material from 30 white, middle-class American women, ages 26–57, who had suffered severe losses of income as a result of divorce. Newman discovered and labeled two groups of women, according to her informants' own accounts of which period in their lives had the greatest effect on how they viewed the world. Women whose adolescent and early married years were in the 1960s and early 1970s seemed to be very different from “women of the Depression” who were born between 1930 and 1940.

These women had grown up in two very different socioeconomic and political environments; the differences in those environments had a profound effect on the shaping of people's subjective, interpretive, and symbolic views of the world, and, according to Newman's analysis, this accounted for differences in how her informants responded to the economic loss of divorce. Newman illustrated her analytic finding with quotes from her informants.

One woman said:

I grew up in the '30s on a farm in Minnesota, but my family lost the farm during the Depression. Dad became a mechanic for the WPA, after that, but we moved around a lot. I remember that we never had any fresh fruits or vegetables during that whole time. At school there were soup lines and food handouts. . . . You know, I've been there. I've seen some hard times and it wasn't pleasant. Sometimes when I get low on money now, I get very nervous remembering those times.

By contrast, “women of the '60s” felt the economic loss of divorce but tended to stress the value of having to be more self-reliant, and the importance of friends, education, and personal autonomy over dependence on material things. Newman illustrated this sentiment with quotes like the following:

Money destroyed my marriage. All my husband wanted was to accumulate more real estate. We had no emotional relationship. Everything was bent toward things. Money to me now is this ugly thing.

Newman found differences in the way women in the two age cohorts dealt with kin support after divorce, the way they related to men in general, and a number of other things that emerged as patterns in her data. For each observation of a patterned difference in response to life after divorce, Newman used selected quotes from her informants to make the point.

Here's another example, from the study I did with Ashton-Vouyoucalos (1976) on Greek labor migrants. Everyone in the population we were studying

had spent 5 years or more in West Germany and had returned to Greece to reestablish their lives. We were interested in how these returned migrants felt about the Greece they returned to, compared with the Germany they left.

Before doing a survey, however, we collected life histories from 15 people, selected because of their range of experiences. Those 15 returned migrants were certainly no random sample, but the consistency of their volunteered observations of differences between the two cultures was striking. Once we noticed the pattern emerging, we laid out the data in tabular form, as shown in table 17.5. The survey instrument that we eventually built reflected the concerns of our informants.

In reporting our findings, Ashton-Vouyoucalos and I referred to the summary table and illustrated each component with selected quotes from our informants. The issue of gossip, for example (under “negative aspects of Greece” in table 17.5), was addressed by Despina, a 28-year-old woman from Thrace. Despina was happy to be back in Greece, but she said:

Look, here you have a friend you visit. Sooner or later you’ll wear or do something she doesn’t like. We have this habit of gossiping. She’ll gossip behind your back. Even if it’s your sister. In Germany, they don’t have that, at least. Not about what you wear or what you eat. Nothing like that. That’s what I liked.

By the way, the translation of Despina’s comment has been doctored to make it sound a bit more seamless than it did in the original. I’ve seen thousands of really interesting quotes in ethnographic reports, and common sense says that most of them were fixed up a bit. I don’t see anything wrong with this. In fact, I’m grateful to writers who do it. Unexpurgated speech is terrible to read. It’s full of false starts, run-ons, fragments, pauses, filler syllables (like “uh” and “y’know”), and whole sentences whose sole purpose is to give speakers a second or two while they think of what to say next. If you don’t edit that stuff, you’ll bore your readers to death.

There is a case to be made for *recording* people’s statements verbatim, but that’s different from *reporting* those statements. Obviously, if you are doing linguistic, conversation, or narrative analysis, you need to record *and* report verbatim material.

Content Analysis

While grounded theory is concerned with the *discovery of hypotheses* from texts and the building of explanatory models from the same or subsequently collected texts, content analysis is concerned with *testing hypotheses* from the

TABLE 17.5
Summary of Repatriates' Ambivalent Statements about Greece

Negative aspects of Greece

Economic

1. Wages are low.
2. Few jobs are available, especially for persons with specialized skills.
3. Working conditions are poor.
4. Inflation is high, especially in the prices of imported goods.

Sociocultural

1. People in general (but especially public servants) are abrupt and rude.
2. The roads are covered with rubbish.
3. Everyone, even friends and relatives, gossips about each other and tries to keep each other down.
4. People of the opposite sex cannot interact easily and comfortably.

Political

1. The government is insecure and might collapse with ensuing chaos or a return to dictatorship.
2. Fear of actual war with Turkey creates a climate of insecurity.

Negative aspects of Germany

Economic

1. Economic opportunities are limited because a foreigner cannot easily open up a private business.
2. People are reluctant to rent good housing at decent prices to migrant workers.

Sociocultural

1. One feels in exile from one's home and kin.
2. Life is limited to house and factory.
3. The weather seems bitterly cold and this furthers the sense of isolation.
4. Migrants are viewed as second-class citizens.
5. Children may be left behind in Greece, to the sometimes inadequate care of grandparents.
6. Lack of fluency in German puts Greek workers at a disadvantage.
7. Parents must eventually choose between sending their children to German schools (where they will grow away from their parents) or to inadequate Greek schools in German cities.
8. Factory routines are rigid, monotonous, and inhuman and sometimes the machinery is dangerous.

Political

1. Migrants have no political voice in Germany or in their home country while they are abroad.
-

SOURCE: "Return Migration to Greece" by H. R. Bernard and S. Ashton-Vouyoucalos, 1976, *Journal of the Steward Anthropological Society* 8:31-51. Table reproduced with permission from the *Journal of the Steward Anthropological Society*.

start—usually, but not always, quantitatively. This requires (1) starting with a theory that you want to test; (2) creating a set of codes for variables in the theory; (3) applying those codes systematically to a set of texts; (4) testing the reliability of coders when more than one applies the codes to a set of texts; (5) creating a unit-of-analysis-by-variable matrix from the texts and codes; and (6) analyzing that matrix statistically with methods like those laid out in chapters 19, 20, and 21.

Content analysis doesn't have to be complicated to be effective. When the Nazis came to power in the 1930s, the U.S. Government Communications Commission began monitoring shortwave radio broadcasts from Germany. Analysts established 14 major propaganda themes in the Nazi media. In 1942, the U.S. Department of Justice accused William Dudley Pelley of sedition, claiming that Pelley was publishing pro-Nazi propaganda while the United States was at war with Germany.

The government asked independent coders to classify 1,240 items in Pelley's publications as belonging or not belonging to one of those 14 Nazi propaganda themes. Harold Lasswell, a political scientist and expert in propaganda analysis, testified that 1,195 of the items (96.4%) "were consistent with and suggested copying from the German propaganda themes" (*United States v. Pelley* 1942). Pelley was convicted. The conviction was upheld by the U.S. Circuit Court of Appeals, and the admissibility in court of evidence based on this simple method of content analysis was established (Goldsen 1947).

Maxine Margolis (1984) did ethnohistorical research on the changing images of women in the United States. She used the *Ladies Home Journal*, from 1889 to 1980, as an archival database, and asked a simple question: Do ads in the *Ladies Home Journal* for household products show homemakers or servants using those products?

From historical data, Margolis knew that the large pool of cheap servant labor in U.S. cities—labor that had been driven there by the Industrial Revolution—was in decline by about 1900. The readers of the *Ladies Home Journal* in those days were middle-class women who were accustomed to employing household servants. Margolis's counts showed clearly the transformation of the middle-class homemaker from an employer of servants to a direct user of household products.

Margolis took a random sample of her database (2 years per decade of the magazine, and 2 months per year, for a total of 36 magazines), but she did not have to devise a complex tagging scheme. She simply looked for the presence or absence of a single, major message. It is very unlikely that Margolis could have made a mistake in coding the ads she examined. Servants are either portrayed in the ad, or they aren't. So, by defining a nominal variable, and one that is easily recognized, Margolis was able to do a content analysis that added

an interesting dimension to her historical ethnographic work on changing images of middle-class urban women.

“Texts” don’t have to be made of words for content analysis. The Codex Borgia is one of just a few surviving books from ancient Mesoamerica. Written in pictographs, it documents, in beautiful, gory detail the ceremonies and gods associated with each day of the 260 days in the ancient Mesoamerican ritual calendar. There were three great civilizations in ancient Mexico: one centered in Teotihuacán, in the central valley; one in Oaxaca, centered in Monte Alban; and the Mayan civilization, which occupied a huge area from southern Mexico to Honduras. The Codex Borgia was painted in Mixtec style—that is, the style from Oaxaca—but the iconography of the Borgia is widely regarded as Nahua—that is, from the central valley. Which raises the question: Who painted the Borgia? It couldn’t have been a Maya, but was it a Nahua or a Mixtec?

John Paddock (1985) noticed that while the focus of the Codex Borgia is on the gods and their associated ceremonies, that of another Mixtec document, the Codex Nuttall, is on events associated with noble families. But both codices have something in common: really gory depictions of human sacrifice—the kind that involved bending the victim over a large, round rock, so that his feet and head were lower than his chest, and then extracting his heart with a flint blade (Paddock 1985:362).

Paddock coded and counted. There are 126 scenes of sacrifice per 100 pages in the Borgia vs. 19 per 100 in the Nuttall. In fact, says Paddock, “Borgia *drags* in blood where it is only remotely appropriate” (ibid.:66) and appears to celebrate human sacrifice, while the other Mixtec codices seem to deplore it. What’s going on?

The Nahua-speaking peoples (the last of whom were the Aztecs) probably came to the Valley of Mexico around 600 A.D., toward the end of the Teotihuacán civilization. Comparing the art of Teotihuacán and that of the Mayans, Paddock notes the absence in Teotihuacán period art, anywhere in Mexico, of violent military scenes. The spread of the Teotihuacán civilization appears to have been the result more of trade than of conquest.

Tezcatlipoca is a fearsome, bloodthirsty god, and a central deity in Borgia, but he doesn’t appear in Oaxaca until after the Aztec invasion in the mid-15th century C.E. Paddock argues that the patron of the Codex Borgia was a Nahua high priest at a large, wealthy center who had a Mixtec painter in his employ and who told the Mixtec painter what to put into the codex. “If the painter was too dainty with the blood,” says Paddock, “he would be corrected. If he showed it too prominently for the patron’s taste, he would hear about that” (Paddock 1985:378). And if he knew what was good for him—if he wanted to stay alive—he did exactly what his patron asked of him.

Content analysis is easily applied to film, Cowan and O'Brien (1990) wanted to know, for example, whether men or women in slasher films were more likely to be survivors, and what other personal characteristics accounted for those who got axed and those who lived. The corpus of text in this case was 56 slasher movies.

These movies contained a total of 474 victims, who were coded for gender and survival. Conventional wisdom about slasher films, of course, holds that victims are mostly women and slashers are mostly men. Although slashers in these films were, in fact, mostly men, it turned out that victims were equally likely to be women or men. Surviving as a female slasher victim, however, was strongly associated with the absence of sexual behavior and with being less physically attractive than nonsurviving women. The male nonsurvivors were cynical, egotistical, and dictatorial. Cowan and O'Brien conclude that, in slasher films, sexually pure women survive and that "unmitigated masculinity" ends in death (*ibid.*:195).

The methodological issues associated with content analysis are all evident here. Does the sample of 56 films used by Cowan and O'Brien justify generalizing to slasher films in general? Did the coders who worked on the project make correct judgments in deciding things like the physical attractiveness of female victims or the personality and behavioral characteristics of the male victims? These two issues in particular, sampling and coding, are at the heart of content analysis. (For content analysis of young adults' reactions to slasher films, see Nolan and Ryan 2000.)

Sampling in Content Analysis

There are two components to sampling in content analysis. The first is identifying the *corpus* of texts; the second is identifying the units of analysis *within* the texts. If you collect 40 or 50 life histories, then you naturally analyze the whole corpus. But when the units of data run into the hundreds or even thousands—like all television commercials that ran during prime time in August 2005; all front-page stories of the *New York Times* from 1851 to 2005; all campaign speeches by John Kerry and George W. Bush during the 2004 presidential campaign—then a representative sample of records must be made.

Gilly (1988) did a cross-cultural study of gender roles in advertising. She videotaped a sample of 12 hours of programming in Los Angeles (United States), Monterrey (Mexico), and Brisbane (Australia), from 8 A.M. to 4 P.M. on Tuesday and from 7 P.M. to 11 P.M. on Wednesday. To control for seasonal variation between the hemispheres, the U.S. and Mexico samples were taken in September 1984, and the Australia sample was taken in February 1985.

There were 617 commercials: 275 from the United States, 204 from Mexico, and 138 from Australia.

Because of her research question, Gilly used only adult men and women who were on camera for at least 3 seconds or who had at least one line of dialogue. There were 169 women and 132 men in the U.S. ads; 120 women and 102 men in the Mexican ads; and 52 women and 49 men in the Australian ads.

Text analysis—particularly nonquantitative analysis—is often based on purposive sampling. Trost (1986) thought the relationship between teenagers and their families might be affected by five different dichotomous variables. To test this idea, he intentionally selected five cases from each of the 32 possible combinations of the five variables and conducted 160 interviews.

Nonquantitative studies in content analysis may also be based on extreme or deviant cases, cases that illustrate maximum variety on variables, cases that are somehow typical of a phenomenon, or cases that confirm or disconfirm a hypothesis. Even a single case may be enough to display something of substantive importance, but Morse (1994) suggests using at least six participants in studies where you're trying to understand the essence of experience and carrying out 30–50 interviews for ethnographies and grounded theory studies.

Once a sample of texts is established, the next step is to identify the basic, nonoverlapping units of analysis. This is called **unitizing** (Krippendorf 1980) or **segmenting** (Tesch 1990). The units may be entire texts (books, interviews, responses to an open-ended question on a survey) or segments (words, word-senses, sentences, themes, paragraphs). If you want to compare across texts—to see *whether or not* certain themes occur—the whole text (representing a respondent or an organization) is the appropriate unit of analysis.

When the idea is to compare the *number of times a theme occurs* across a set of texts, then you need to break the text down into smaller chunks—what Kortendick (1996) calls **context units**—each of which reflects a theme.

Coding in Content Analysis

With a set of texts in hand, the next steps are to develop a codebook and actually code the text. Up to this point, this is what's done in grounded-theory analysis. Consider Elizabeth Hirschman's work (1987) on how people sell themselves to one another in personal ads. From her reading of the literature on resource theory, Hirschman thought that she would find 10 kinds of resources in personal ads: love, physical characteristics, educational status, intellectual status, occupational status, entertainment services (nonsexual), money status, demographic information (age, marital status, residence), ethnic

characteristics, and personality info (not including sexual or emotional characteristics).

Hirschman formulated and tested specific hypotheses about which resources men and women would offer and seek in personal ads. She selected 20 test ads at random from the *New York Magazine* and *The Washingtonian* and checked that the 10 kinds of resources were, in fact, observable in the ads. Sexual traits and services were less than 1% of all resources coded. This was 1983–1984, but even then, ads with explicit references to sexual traits and services were more common in other periodicals than in *The Washingtonian* and *New York Magazine*.

Hirschman next gave 10 men and 11 women the list of resource categories and a list of 100 actual resources (“young,” “attractive,” “fun loving,” “divorced,” “32-year-old,” etc.) gleaned from the 20 test ads. She asked the 21 respondents to match the 100 resources with the resource category that seemed most appropriate. This exercise demonstrated that the resource items were **exhaustive** and **mutually exclusive**: No resource items were left over, and all of them could be categorized into only one of the 10 resource categories.

When she was confident her codebook worked, Hirschman tested her hypotheses. She sampled approximately 100 female-placed ads and 100 male-placed ads from each magazine—a total of 400 ads. A male and a female coder, working independently (and unaware of the hypotheses of the study), coded 3,782 resource items taken from the 400 ads as belonging to one of the 10 resource categories. The coding took 3 weeks. This is not easy work.

Hirschman was concerned with **intercoder reliability**—that is, making sure that coders saw the same thing when they coded those ads. She gave the data to a third coder who identified discrepancies between the first two coders. Of 3,782 resource items coded, there were discrepancies (theme contrasts) on 636 (16.8%), and one of the coders failed to code 480 items (12.7%). Hirschman resolved the theme contrasts herself. She checked the omissions against the ads to see if the coder who had made an assignment had done so because the resource was, in fact, in the ad. This was always the case, so the 480 resource items omitted by one coder were counted as if they had been assigned to the ad by both coders.

The results? Men were more likely than women to *offer* monetary resources; women were more likely than men to *seek* monetary resources. Women were more likely than men to offer physical attractiveness. Washington, D.C., and New York City are supposed to be hip places, yet the way men and women wrote their own personal ads in 1983–1984 conformed utterly to traditional gender role expectations. Other studies since then show that, while there are some differences across social and economic classes, what hetero-

sexual men and women offer each other in personal ads in the United States hasn't changed much since 1983 (Wiederman 1993; Willis and Carlson 1993; Goode 1996; Butler-Smith et al. 1998; Cameron and Collins 1998). Of course, we expect that people of different sexual orientations and people in different cultures and subcultures will seek and offer different resources in personal ads, which sets up a program of work on cross-cultural content analysis. (See de Sousa Campos et al. [2002] for a content analysis of personal ads in Brazil; see Parekh and Beresin [2001] for an analysis of ads in the United States, India, and China. And see Yancey and Yancey [1997] for a content analysis of personal ads of people seeking interracial relationships in the United States. See Smith and Stillman 2002 on differences in the personal ads of heterosexual, lesbian, and bisexual women; and see Kaufman and Phua 2003 on differences in the personal ads of gay and straight men.)

Intercoder Reliability

It is quite common in content analysis to have more than one coder mark up a set of texts. The idea is to see whether the constructs being investigated are shared—whether multiple coders reckon that the same constructs apply to the same chunks of text. There is a simple way to measure agreement between a pair of coders: you just line up their codes and calculate the percentage of agreement. This is shown in table 17.6 for two coders who have coded 10 texts for a single theme, using a binary code, 1 or 0.

TABLE 17.6
Measuring Simple Agreement between Two Coders on a Single Theme

	<i>Units of Analysis (Documents/Observations)</i>									
	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>	<i>10</i>
Coder 1	0	1	0	0	0	0	0	0	1	0
Coder 2	0	1	1	0	0	1	0	1	0	0

Both coders have a 0 for texts 1, 4, 5, 7, and 10, and both coders have a 1 for text 2. These two coders agree a total of six times out of ten—five times that the theme, whatever it is, does not appear in the texts, and one time that the theme does appear. On four out of ten texts, the coders disagree. On text 9, for example, Coder 1 saw the theme in the text, but Coder 2 didn't. Overall, these two coders agree 60% of the time.

The total observed agreement, though, is not a good measure of intercoder reliability because people can agree that a theme is present or absent in a text

just by chance. To adjust for this possibility, many researchers use a statistic called **Cohen’s kappa** (Cohen 1960), or *k*.

Cohen’s kappa

Kappa is a statistic that measures *how much better than chance* is the agreement between a pair of coders on the presence or absence of binary (yes/no) themes in texts. Here is the formula for kappa:

$$k = \frac{\text{Observed} - \text{Chance}}{1 - \text{Chance}} \qquad \textbf{Formula 17.1}$$

When *k* is 1.0, there is perfect agreement between coders. When *k* is zero, agreement is what might be expected by chance. When *k* is negative, the observed level of agreement is less than what you’d expect by chance. And when *k* is positive, the observed level of agreement is greater than what you’d expect by chance. Table 17.7 shows the data in table 17.6 rearranged so that we can calculate kappa.

TABLE 17.7
The Coder-by-Coder Agreement Matrix for the Data in Table 17.6

		<i>Coder 2</i>		<i>Coder 1 totals</i>
		<i>Yes</i>	<i>No</i>	
<i>Coder 1</i>	<i>Yes</i>	1 (<i>a</i>)	1 (<i>b</i>)	2
	<i>No</i>	3 (<i>c</i>)	5 (<i>d</i>)	8
<i>Coder 2 totals</i>		4	6	10 (<i>n</i>)

The *observed agreement* between Coder 1 and Coder 2 is:

$$\frac{a + d}{n}$$

Here, Coder 1 and Coder 2 agreed that the theme was present in the text once (cell *a*) and they agreed that the theme was absent five times (cell *d*), for a total of 6, or 60% of the 10 texts.

The probability that Coder 1 and Coder 2 agree by chance is:

$$\frac{a + b}{n} \times \frac{a + c}{n} + \frac{c + d}{n} \times \frac{b + d}{n}$$

Here, the probability that Coder 1 and Coder 2 agreed by chance is .08 + .48 = .56. Using formula 17.1, we calculate kappa:

$$k = \frac{.6 - .56}{1 - .56} = .0909$$

In other words, the 60% observed agreement between the two coders for the data in table 17.6 is about 9% better than we'd expect by chance. Whether we're talking about agreement between two people who are coding a text or two people who are coding behavior in a time allocation study, 9% better than chance is nothing to write home about.

Carey et al. (1996) asked 51 newly arrived Vietnamese refugees in New York State 32 open-ended questions about tuberculosis. Topics included knowledge and beliefs about TB symptoms and causes, as well as beliefs about susceptibility to the disease, prognosis for those who contract the disease, skin-testing procedures, and prevention and treatment methods. The researchers read the responses and built a code list based simply on their own judgment. The initial codebook contained 171 codes.

Then, Carey et al. broke the text into 1,632 segments. Each segment was the response by one of the 51 respondents to one of the 32 questions. Two coders independently coded 320 of the segments, marking as many of the themes as they thought appeared in each segment. Segments were counted as reliably coded if both coders used the same codes on it. If one coder left off a code or assigned an additional code, then this was considered a coding disagreement.

On their first try, only 144 (45%) out of 320 responses were coded the same by both coders. The coders discussed their disagreements and found that some of the 171 codes were redundant, some were vaguely defined, and some were not mutually exclusive. In some cases, coders simply had different understandings of what a code meant. When these problems were resolved, a new, streamlined codebook was issued, with only 152 themes, and the coders marked up the data again. This time they were in agreement 88.1% of the time.

To see if this apparently strong agreement was a fluke, Carey et al. tested intercoder reliability with kappa. The coders agreed perfectly ($k = 1.0$) on 126 out of the 152 codes that they'd applied to the 320 sample segments. Only 17 (11.2%) of the codes had final k values = 0.89. As senior investigator, Carey resolved any remaining intercoder discrepancies himself (Carey et al. 1996).

How much intercoder agreement is enough? As with so much in real life, the correct answer, I think, is: it depends. It depends, for example, on the level of inference required. If you have texts from single mothers about their efforts to juggle home and work, it's easier to code for the theme "works full-time" than it is to code for the theme "enjoys her job."

It also depends on what's at stake. X-rays are texts, after all, and I'd like a pretty high level of intercoder agreement if a group of physicians were deciding on whether a particular anomaly meant my going in for surgery or not. In text analysis, the standards are still evolving. Many researchers are satisfied with kappa values of around .70; others like to shoot for .80 and higher (Krippendorf 1980; Gottschalk and Bechtel 1993).

HRAF: Cross-Cultural Content Analysis

In the 1940s, George Peter Murdock, Clellan S. Ford, and other behavioral scientists at Yale led the effort to organize an interuniversity, nonprofit organization that is now the Human Relations Area Files (HRAF) at Yale University. HRAF is the world's largest archive of ethnography, with about a million pages of text, collected from more than 7,800 books and articles, on over 400 cultural groups around the world. The archive is growing at about 40,000 pages a year, and about 40% of the material is now available around the world via the Internet through libraries at institutions that subscribe. (Go to: <http://www.yale.edu/hraf/>.)

Pages of the HRAF database are indexed by professional anthropologists, following the *Outline of Cultural Materials*, or OCM. This is a massive indexing system that was developed by Murdock and others (2004 [1961]) to organize and classify material about cultures and societies of the world. The OCM is used by cross-cultural researchers to find ethnographic data for testing hypotheses about human behavior across cultures. (Some anthropologists use it to code their field notes; see chapter 14.)

There are 82 main domains in the OCM, in blocks of 10, from 10 to 91. Block 16, for example, is about demography. Within this block there are eight subdomains labeled 1611, 1622, . . . 168. These domains cover specific topics like mortality (code 165), external migration (code 167), and so on. Block 58 covers the family with codes for nuptials (585), termination of marriage (586), etc. Other major blocks of codes are for domains like kinship, entertainment, social stratification, war, health and welfare, sickness, sex, religious practices. . . .

HRAF turns the ethnographic literature into a database for content analysis and cross-cultural tests of hypotheses because you can search the archive for every reference to any of the codes across the more than 400 cultures that are covered.

Doing Cross-Cultural Text-Based Research

There are five steps in doing an HRAF study (Otterbein 1969):

1. State a hypothesis that requires cross-cultural data.
2. Draw a representative sample of the world's cultures.
3. Find the appropriate OCM codes in the sample.
4. Code the variables according to whatever conceptual scheme you've developed in forming your hypothesis.
5. Run the appropriate statistical tests and see if your hypothesis is confirmed.

Sampling and Galton's Problem

In 1889, Edward Tylor gave a paper at the Royal Society in London in which he tried to relate, among other things, marital residence (matrilocal, patrilocal, etc.) to customs of kin avoidance. Francis Galton, asked: "Weren't some of the societies that Mr. Tylor was using in his analysis related to one another? Wouldn't that negate using them each as an independent example of the same variable?"

This became known as **Galton's Problem**. One way to deal with it is to use the Standard Cross-Cultural Sample of 186 societies (Murdock and White 1969) or the HRAF Probability Sample of 60 societies (HRAF 1967) as your sample. These samples were developed so that the societies would be independent of one another, linguistically and culturally.

Another way is to treat the relation between pairs of societies as an independent variable—that is, to measure how close they are linguistically, culturally, or geographically—and to use that variable as a predictor of whatever hypothesis you're testing (Dow et al. 1984).

Or you can, as Carol and Melvin Ember suggest (2001:89), choose a simple random sample from the societies that are currently described in the HRAF archives. If you choose 60 societies samples at random, you can run your test on 30 of them and then compare your results to the other 30. If you get the same answers twice, you can be much more confident about them than if you get them just once.

Coding

If a piece of an ethnography is coded as 682, this means that it is about offenses against life. It does not tell you what the offense is. It may be killing someone by accidentally setting fire to their house; it may be paying a sorcerer to cast a deadly spell on someone who subsequently dies; it may be a revenge or honor killing; or it may be an inappropriate act in battle. The only way to

tell is to read the material and convert the primary ethnographic material into usable codes for statistical analysis. You might use 1 = war, 2 = rape, 3 = use of witchcraft, and so on, depending on your particular research problem.

In a classic study, Landauer and Whiting (1964) tested the association between stress (both physical and emotional) in infancy and height. Psychologists had found in lab experiments that stroking rat pups (a very stressful thing to do to a baby rat) led to longer adult rats. Do humans respond as rats do to physical and emotional stress in infancy? Landauer and Whiting tested this on two samples of societies across the world for which sufficient data were available.

The dependent variable here is mean adult male height and the independent variable is the presence of piercing (lips, nose, scarification, circumcision) or molding (of arms or legs or head) during the first 24 months of life. Table 17.8 shows the result.

TABLE 17.8
The Results of Landauer and Whiting's Study (1964)

	<i>Piercing or molding present during first 24 months</i>	
Study 1 (35 societies)	<i>Present (n = 17)</i>	<i>Absent (n = 18)</i>
Mean height	65.18 inches	62.69 inches
	$t = 3.72, p < .002$	
Study 2 (30 societies)	<i>Present (n = 19)</i>	<i>Absent (n = 11)</i>
Mean height	66.05 inches	63.41 inches
	$t = 4.68, p < .001$	

Landauer and Whiting knew they had to be very cautious about interpreting the results in table 17.8 because correlation does not, by itself, imply cause. It could be that parents who put their infants through this kind of stress give those children more food or better medical care, which supports growth. Or it could be that boys who are stressed during infancy become more aggressive and only the tallest survive. "This," said Landauer and Whiting, "is the problem with correlational research" (1964:1018).

Still, they were able to control for, and rule out, variations in sunlight (and hence in the body's production of vitamin D) and variations in population genetics—two factors that are well known to cause variations in height. More important, they took an idea from the lab (where internal validity is strong because of the controlled conditions of the experiment) and tested it in the field. Yes, the experiments were on lab animals, but the experiments on rats

were about manipulation (all that stroking) of endocrines that are known to control growth. Why not, they asked, test this across species, as well as across cultures?

The key to this research is not the data about adult height. It's the coding of the ethnographic, textual data about piercing, molding, scarring, circumcising, and so on. In the end, in content analysis as in all research, *you* have to make the measurements.

Over the years, as researchers have used the Human Relations Area Files, they've read through the primary materials and coded variables that were germane to their particular studies. Barry and Schlegel (1980) edited a book containing published codes for several hundred variables on the 186-society sample developed by Murdock and White (1969). The *World Cultures Journal* (WCJ) has published most of the codes in Barry and Schlegel on disk and continues to publish codes from cross-cultural studies (<http://eclectic.ss.uci.edu/~drwhite/worldcul/world7.htm>). Two journals, *Ethnology* and *Cross-Cultural Research* (formerly *Behavior Science Research*), continue to publish the variables and codes used by cross-cultural researchers who publish papers in those journals and codes for 300 variables have been published on disk by HRAF Press for a 60-society sample.

Coding is painstaking work. Carol and Melvin Ember (2002) were interested in the relationship between various aspects of child rearing (If corporal punishment is present in a society, is it usually the mother or the father who does it? Are fathers usually present or absent in infants' lives? How much does the father care for infants?) and the amount of interpersonal violence in societies by adult men. Some of the codes they used on child-rearing practices had been published elsewhere, but it took the Embers 4 years to code the variables for frequency of war and interpersonal aggression in the 186-society sample, so, like most cross-cultural researchers, they published those codes (Ember and Ember 1992).

Testing hypotheses on published codes contributes to theory in anthropology. Using published codes, Barber (1998) found that the frequency of male homosexual activity was low in hunting and gathering societies and increased with the complexity of agricultural production. Ethnographic reports of male homosexuality were also more likely for societies in which women did not control their own sexuality—a well-known correlate of increased reliance on complex agriculture.

Other Problems

Information is not always where you expect it to be in the Human Relations Area Files. There is no code in the OCM for family violence, so when David

Levinson (1989) did the research for his book on this topic, he asked the coders at HRAF how *they* would code it. They said that they would classify family violence under code 593 (family relations) or code 578 (in-group antagonisms).

Levinson scoured the HRAF for references to those codes, and did, in fact, find a lot of material. He coded whether or not a society was reported to exhibit family violence, what kind of violence was reported (child abuse, abuse of the elderly, etc.), and how severe the violence was. Later, while he was browsing through the files, Levinson noticed that wife beating—clearly, a form of family violence—was usually coded under 684, sex and marital offenses. Wife beating, it turns out, is reported for many societies only in cases of adultery or suspicion of adultery, and hence, under code 684 (Levinson, personal communication). The lesson for conducting cross-cultural research is pretty clear: There is no substitute for reading the ethnographies and looking for new clues on how to code variables.

There is also no substitute for good data. The quality of ethnographic data depends crucially on the skill of the person who collected and recorded those data. Ethnographers may be biased in recording their data. Even if they are not biased, they may have the misfortune to interview incompetent informants—or even people who just lied. Raoul Naroll discussed these and other problems in a pioneering work in 1962. Since then, cross-cultural researchers have done many studies on data quality control (see Rohner et al. 1973; Levinson 1978).

William Divale (1976), for example, tested the long-standing idea that women have higher status as societal complexity increases. He correlated two independent measures of female status with societal complexity across a representative sample of cultures and found—against all expectations—that the higher the complexity of the society, the *lower* the status of women.

Divale recalculated the correlation using only ethnographies written by people who had spent at least a year in the field and who spoke the native language fluently. When he did this, he found high female status reported at all levels of societal complexity, with low status reported primarily among less-complex societies. The unexpected relation between female status and societal complexity vanished.

Computers and Text Analysis

Anyone who collects mountains of text will want to take advantage of modern text analysis software. Don't take the phrase "text analysis software" literally. Computer programs do a lot, but in the end, *you* do the analysis; *you*

make the connections and formulate hypotheses to test; *you* draw conclusions and point them out to your readers.

The two broad approaches in text analysis—content analysis and grounded theory—are reflected in the available software. Programs for automated content analysis are based on the concept of a computerized, contextual dictionary. You feed the program a piece of text; the program looks up each word in the dictionary and runs through a series of disambiguation rules to see what the words mean (for example, whether the word “concrete” is a noun (stuff you pave your driveway with) or an adjective (“I need a concrete example of this or I won’t believe it”).

In the 1960s, Philip Stone and others (1966) developed a program for doing automated content analysis called the *General Inquirer*. They tested it on 66 suicide notes—33 written by men who had actually taken their own lives, and 33 written by men who were asked to produce simulated suicide notes. The program parsed the texts and picked the actual suicide notes 91% of the time (Ogilvie et al. 1966). The latest version of the system (which runs with a dictionary called the Harvard IV-4) has a 13,000-word dictionary and over 6,000 rules. It can tell whether the word “broke” means “fractured,” or “destitute,” or “stopped functioning,” or (when paired with “out”) “escaped” (Rosenberg et al. 1990:303). (For more about the *General Inquirer*, see appendix F.)

About the same time, Nicholas Colby (1966) developed a special-purpose dictionary for Zuni and Navajo texts. From his ethnographic work, Colby had the impression that the Navajo regarded their homes as havens and places of relaxation but the Zuni viewed their homes as places of discord and tension. To test this idea, Colby developed two groups of words, one group associated with relaxation (words like assist, comfort, affection, happy, and play) and one associated with tension (words like discomfort, difficult, sad, battle, and anger). He then had the computer look at the 35 sentences that contained the word “home” *and* one of the words in the two word groups. Navajos were more than twice as likely as the Zuni to use relaxation words when talking about home than they were to use tension words; the Zuni were almost twice as likely as the Navajo to use tension words when they talked about their home.

Over the years, computer-assisted content analysis has developed into a major industry. When you hear “this call may be monitored for quality assurance purposes,” it’s likely that the conversation will be turned into text that will be submitted to a high-end data-mining program for analysis.

Most content analysis, however, is not based on computerized dictionaries. It’s based on the tried-and-true method of coding a set of texts for themes, producing a text-by-theme profile matrix, and then analyzing that matrix with statistical tools. Most text analysis packages today support this kind of work.

You code themes on the fly, as you read the text on the screen, and the program produces the text-by-theme matrix for you. Then, you import the matrix into your favorite stats package (SPSS®, SAS®, SYSTAT®, etc., see appendix F) and run all the appropriate tests (about which, more in the chapters on quantitative analysis).

EZ-Text, from the U.S. Centers for Disease Control and Prevention, does this kind of work. It also lets you search for segments of text by themes so that you can study the relation of themes and/or use selected quotes from your informants in your write-up. EZ-Text is available free from the U.S. Centers for Disease Control and Prevention, as is AnSWR, the companion program for large, multi-researcher projects.

EZ-Text and AnSWR support grounded theory research, but commercial programs for this kind of work offer a lot more. Programs like Atlas/ti and NVivo let you code and retrieve text, create matrices, and build models of how the themes in your data are associated with each other. (See Lewis [2004] for a review of these two widely used packages.) Some programs, like Atlas/ti, let you code audio, still images, and video files directly, so you can treat all kinds of qualitative data as text. And some programs, like QDA Miner and C-I-SAID, are designed to do both quantitative content analysis and qualitative, grounded-theory research. (For URLs to the sites for all these programs, see appendix F.)

Programs for analyzing qualitative data are getting more and more sophisticated, with more features added all the time. Were transcription not such a nuisance, social science would have focused long ago on the wealth of qualitative data that describe life across the globe. But with voice-recognition software coming on strong, and transcription becoming less and less intimidating, all forms of text analysis—narrative analysis, discourse analysis, grounded theory, content analysis—will become more and more attractive.

Eventually, I expect text analysis software will have built-in modules for doing everything I've covered in this chapter—including kappa and other tests for assessing intercoder reliability—as well as many other qualitative methods, like those covered in the next chapter: ethnographic decision modeling, Boolean analysis, taxonomies, and componential analysis. Text analysis is only just beginning to come into its own. It's going to be very exciting.



Qualitative Data Analysis II: Models and Matrices

In chapter 17, I focused on the analysis of texts—analyses that require identifying and organizing themes. I focus in this chapter on an altogether different set of qualitative analyses: **ethnographic decision models**, **folk taxonomies** or **taxonomic analysis**, **componential analysis**, and **analytic induction**. These methods all have one thing in common: They are based on some principles of systematic logic. More about this as we move along. Let's get right to the first method: ethnographic decision modeling.

Ethnographic Decision Models (EDMs)

Ethnographic decision models predict the choices that people will make under specific circumstances. There is absolutely nothing quantitative about ethnographic decision models (EDMs), but they are highly formal and it takes some practice to get the hang of them. They have been used to study how fishermen decide where to fish (Gatewood 1983b) or what price to place on their products (Gladwin 1971; Quinn 1978), and how people decide on which treatment to use for an illness (Young and Garro 1994 [1981]).

Here's the idea behind decision models. Suppose I could ask a farmer some questions, none of which is "What did you plant last year?" When I'm done, I make a prediction about what the farmer planted. *Then* I ask the farmer what crop he or she planted and see if I got it right. Christina Gladwin has modeled these decisions among farmers around the world—in Malawi, Guatemala, the United States, Mexico, Peru and gets it right between 80% and 90% of the time (Gladwin 1976, 1980, 1983, 1989, and personal communication).

Predicting a dependent variable (like planting corn instead of some other crop) from some independent variables (like a farmer's financial condition or family size) is the goal of statistical analysis, which is the subject of the next few chapters. But ethnographic decision modeling is based on asking questions, sorting out some logical rules about how the questions have to be ordered, and laying out the order in a picture (like a tree diagram) or in writing. It is, in other words, entirely qualitative. As with all cognitive research methods, we don't know if EDMs just predict behavior or if they also reflect the way people think about things. The jury is still out on that one. But getting it right 80 to 90% of the time is as good as it gets in the social sciences. Ethnographic decision modeling is a qualitative method with real power.

How to Build EDMs

Gladwin (1989) lays out the method for building an ethnographic decision tree model. First, of course, you have to decide on the decision you are studying and what the alternatives are in that decision. I'll use the decision "to make your 8 A.M. class or not" as an example. The alternatives are yes and no.

A grand tour ethnographic question like "Tell me about why people go to class or skip 8 A.M. classes" will get you a lot of information about the alternatives and the reasons for the alternatives, especially from expert informants. The major alternatives are: Get up and go to class, get up and do something else, sleep in. The "get up and do something else" alternative consists of a list: lounge around, watch old soaps on the tube, study for an exam later in the day, and so on.

To make your ethnographic knowledge about the decision more formal—that is, to build an EDM—track down Alex, a respondent who has an 8 A.M. class and ask: "Did you make your 8 A.M. class today?" When he answers, ask him: "Why [did you] [didn't you] go to that class?" Suppose he says "I went to class today because I *always* go to class unless I'm sick." Ask him: "Were you sick this morning?" Record his answer and draw a tree diagram (also called a **dendrogram**), like the one in figure 18.1, to represent his decision.

Figure 18.1 accounts perfectly for Alex's decision. It has to; it contains nothing more than the information from the ethnographic interview with Alex.

Now go to your second respondent, Sheila, who says that yes, she went to her 8 A.M. class. Why? "It's a really tough class," she says, "If I miss one of those classes, I'll never catch up."

Every reason for your respondents' decisions becomes a question you can ask. Use what you learned from your interview with Alex and ask Sheila: "Do

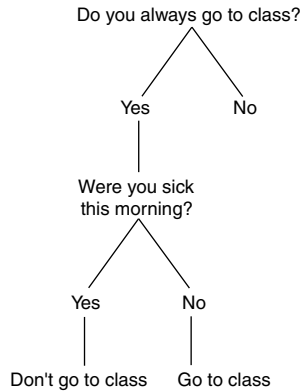


Figure 18.1. An ethnographic decision model after interviewing one informant (Alex).

you *always* go to class?” Sheila says that she sometimes skips early classes if those classes are really easy and she needs to study for an exam in another class later in the day. Ask her: “Were you sick this morning?” If she says, “no,” draw the diagram in figure 18.2.

Your third respondent, Brad, says that no, he didn’t go to class this morning; no, he doesn’t always go to class; yes, he skips class when he’s sick; no, he wasn’t sick this morning; no, he didn’t have an exam later in the day; no, his 8 A.M. class isn’t tough; but he was out very late last night and just didn’t feel like going to class this morning. Figure 18.3 combines all the information we have for Alex, Sheila, and Brad.

In fact, we don’t know if Sheila was out late last night, and if she had been, whether that would have affected her decision to go to class early this morning. We can find out by going back and asking Sheila the new question. We could also go back and ask Alex if he had an exam later in the day and if he’d been out late last night.

But we won’t. In practice, it is very difficult to go back to informants and ask them all the questions you accumulate from EDM interviews. Instead, the usual practice is to build a composite diagram, like the one in figure 18.3, and push on. We also won’t ask Brad what he *would* have done if he’d had a really tough 8:00 A.M. class and had been out late the night before. In building EDMs, *we deal only with people’s reports of their actual, most recent behavior.*

Eventually, you’ll stop getting new decisions, reasons, and constraints. Building a model for the decision on a single campus to attend or not attend

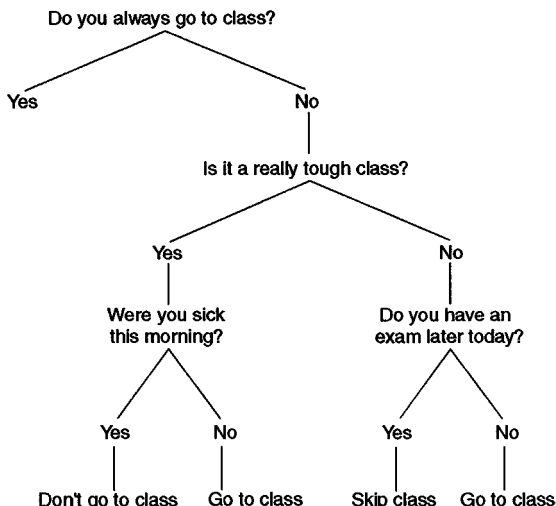


Figure 18.2. An ethnographic decision model after interviewing two informants (Alex and Sheila).

early classes probably won't require more than 20 informants. Building an EDM that accounts for this decision on two very different campuses—say, a small, private college and a huge state school—may take twice that many informants. As the culture gets more heterogeneous, the sample size needed in order to find a stable decision model goes up. Accounting for the decisions of students in New York City and Sussex, England, will require more interviews. Add Mexico City and the number may double again. Add Cameroon. . . .

Testing an EDM

Figure 18.3 may account for all the decisions of your next several informants, but eventually you'll run into an informant who says she doesn't always go to class, she wasn't sick this morning, she doesn't have a tough class at 8 A.M., she wasn't out late last night, and she still didn't make her early class today. Why? Because she just didn't feel like it.

If you add enough constraints, you can always build a model to account for every decision of every respondent. But what's the use of a model with 20 different sets of ordered reasons that accounts for the decisions of 20 informants? You might as well just ask every informant to explain his or her actions. The trick is to model the decisions of, say, 18 out of 20 informants

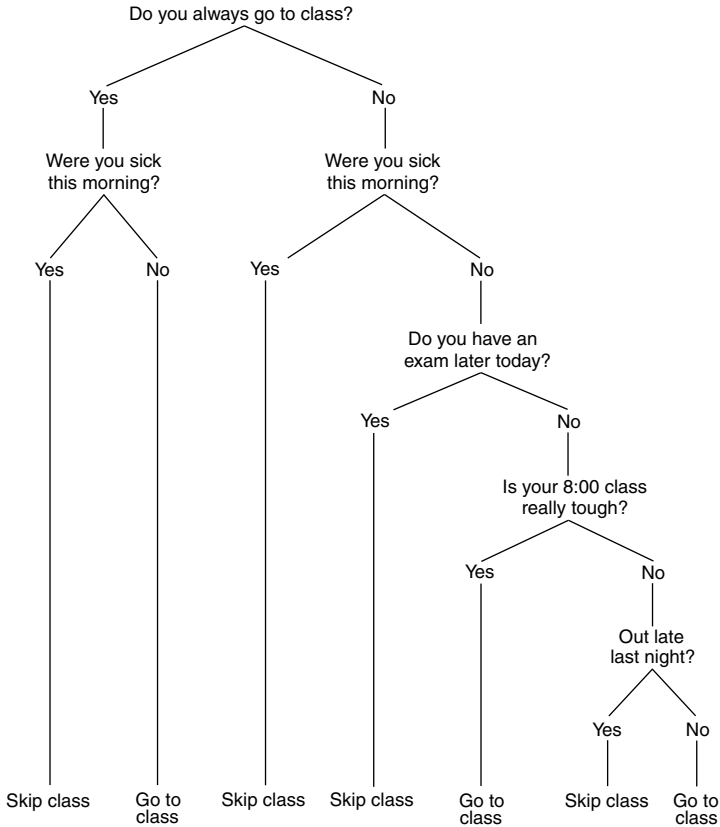


Figure 18.3. An ethnographic decision model after interviewing three informants (Alex, Sheila, and Brad).

(90% prediction) with just a handful of ordered rules. So, when you stop getting new reasons or constraints from EDM interviews, try building a model that accounts for at least 80% of the decisions with the fewest number of rules. Then—and here’s the important part—test your model against an entirely new group of informants.

The interview for the second group of informants is rather different. Now you do ask them all the questions in your model (that is, you probe for all the reasons and constraints for the decision that your first group taught you), and then you guess what their decision was. In our example, you’d interview 20 or 30 new informants, all of whom have 8 A.M. classes, and you’d ask each one: Do you always go to class? Were you sick this morning? Were you out

late last night? and so on, exhausting all the questions from your model. If your model works, you'll be able to *predict* the decisions of the second group from their answers to the model's questions.

Representing Complicated Models with Tables and IF-THEN Charts

Young and Garro's EDM: Decision Tables

James Young and Linda Garro studied how Tarascan people in Pichátaro, Mexico, choose one of four ways to treat an illness: Use a home remedy, go to a native curer, see a *practicante* (a local, nonphysician practitioner of modern medicine), or go to a physician (see Young 1980; Young and Garro 1982, 1994 [1981]; Garro 1986). From their ethnographic work, Young and Garro believed that the decision to use one or another of these treatments depended on four factors:

1. how serious an illness was perceived to be (gravity);
2. whether a home remedy for the illness was known;
3. whether the informant had confidence in the general efficacy of a mode of treatment for a particular illness; and
4. the accessibility (in terms of cost and transportation) of a particular mode of treatment.

The choice situations emerged from structured interviews with eight men and seven women who were asked:

If you or another person in your household were ill, when—for what reasons—would you [consult] [use] _____ instead of [consulting] [using] _____? (Young and Garro 1994 [1981]:132)

Young and Garro used this question frame to elicit responses about all six possible pairs of treatment alternatives: home remedy vs. a physician, curer vs. home remedy, and so on. To check the validity of the statements made in the interviews, Young and Garro collected case histories of actual illnesses and their treatments from each of the 15 informants.

Next, the researchers completed interviews with 20 informants using a series of "What if . . ." questions to generate decisions, under various combinations of circumstances, regarding the selection of treatments for illnesses. For example, informants were asked:

Let's say there is a person who has a very grave illness. In this family, money is scarce—sure, they're eating, but there is just not anything left over. They have

had this illness in the family before, and they now know of the remedy that benefited the illness on the previous occasion. What do you think they are going to do? (ibid.:137)

This vignette combines the condition of a serious illness (level 3 on gravity in tables 18.1 and 18.2), with lack of accessibility (no money), and a known

TABLE 18.1
Young and Garro's Decision Table Showing How Pichtatareños Choose
an Initial Method of Treating an Illness

<i>Rules:</i>	1	2	3	4	5	6	7	8	9
Conditions									
gravity ^a	1	1	1	2	2	2	3	3	3
known home remedy ^b	Y	N	N	Y	N				
faith ^c		F	M	(F)	F	M	F	M	(M)
accessibility ^d								N	Y
Choices									
self-treatment	X			X					
curer		X			X		X		
<i>practicante</i>			X			X		X	
physician									X

a. 1 = nonserious, 2 = moderately serious, 3 = grave

b. Y = yes, N = no

c. F = favors folk treatment, M = favors medical treatment

d. Y = money and transportation available, N = either money or transportation not available

SOURCE: Reprinted by permission of Waveland Press, Inc., from J. C. Young and L. C. Garro, *Medical Choice in a Mexican Village*, 1981 (reissued 1994), p. 154. All rights reserved.

remedy that can be applied at home. Young and Garro used the three levels of gravity, two possible conditions of knowing a remedy (yes and no), and two possible conditions of accessibility (yes and no) in making up the vignettes, which meant that they had to make up eight of them. Each vignette was presented to each informant for a response. Tables 18.1 and 18.2 show the **decision tables** for Young and Garro's data.

From these qualitative data, collected in structured interviews, Young and Garro developed their decision model, for the initial choice of treatment. The model, containing nine decision rules, is shown in table 18.1. Rule number 1, for example, says that if the illness is not serious and there is a known home remedy, then treat the illness yourself. Rule number 9 says that for grave illnesses there is an implicit understanding that physicians are better (hence the M in parentheses), so if there is money, then go to a physician.

Rule number 9 also says that for the few cases of very grave illnesses where

TABLE 18.2
 Young and Garro's Decision Table Showing How Pichatareños Choose a Method of Treating an Illness When Their First Choice Doesn't Work

Rules:	1	2	3	4	5	6	7	8	9	10	11
Conditions											
preceding choice ^a	ST	ST	ST	ST	C-P	C-P	C	P	Dr	Dr	Dr
current gravity ^b		1-2	3	3	1	2-3	2-3	2-3			
faith ^c	F	M	M	(M)							M
accessibility ^d			N	Y		Y	N	N		N	Y
Choices											
self-treatment					X						
curer	X							X	X	X	
<i>practicante</i>		X	X				X				
physician				X		X					X

a. ST = self-treatment, C = curer, P = *practicante*, Dr = physician
 b. 1 = nonserious, 2 = moderately serious, 3 = grave
 c. F = favors folk treatment, M = favors medical treatment
 d. Y = money and transportation available, N = either money or transportation not currently available
 SOURCE: Reprinted by permission of Waveland Press, Inc., from J. C. Young and L. C. Garro, *Medical Choice in a Mexican Village*, 1981 (reissued 1994), p. 154. All rights reserved.

physicians are commonly thought not to be effective, apply rule number 7 and go to a curer. The blank spaces in the top part of table 18.1 indicate irrelevant conditions. In rule number 1, for example, there is no question about accessibility for home remedies because they cost little or nothing and everyone has access to them.

Sometimes the treatment selected for an illness doesn't work and another decision has to be made. Table 18.2, with 11 decision rules, shows Young and Garro's analysis of this second stage of decision making. Their entire two-stage model is based on their sense of emerging patterns in the data they collected about decision making. The question, of course, is: Does it work?

Young and Garro tested their model against 444 treatment choices gathered from 62 households over a 6-month period. To make the test fair, none of the informants in the test were among those whose data were used in developing the model. Table 18.3 shows the results of the test. There were 157 cases covered by rule number 1 from table 18.1 (first-stage decision), and in every single case informants did what the rule predicted. In table 18.3, errors (informants' choices that are not predicted by the model) are in parentheses, so informants did what rule number 6 predicted 20 out of 29 times.

Overall, for the first stage, Young and Garro's decision rules predict about 94% of informants' reported behavior. After removing the cases covered by

TABLE 18.3
 Test Results of Young and Garro's Decision Model of How Pichatareños
 Choose a Treatment Method When They Are Ill

Table	Rule	<i>Treatment Method Chosen</i>				Totals	Percentage Correct
		<i>Self-Treatment</i>	<i>Curer</i>	<i>Practicante</i>	<i>Physician</i>		
18.1	1	157				157	
	2		4			4	
	3			5		5	
	4	67			(1)	68	
	5		8			8	
	6	(2)		20	(7)	29	
	7		8			8	
	8		(2)	4	(2)	8	
	9			(2)	11	13	
				Subtotal	300	94.7%	
18.2	1		19			19	
	2		(1)	28	(6)	35	
	3		(3)	6		9	
	4			(2)	22	24	
	5	3	(1)			4	
	6	(2)	(2)	(1)	24	29	
	7	(1)		3	(2)	6	
	8		2	(1)		3	
	9	(1)	7			8	
	10					0	
	11				7	7	
				Subtotal	144	84.0%	
				Total	444	91.2%	

SOURCE: Reprinted by permission of Waveland Press, Inc., from J. C. Young and L. C. Garro, *Medical Choice in a Mexican Village*, 1981 (reissued 1994), p. 154. All rights reserved.

rules 1 and 4 (which account for half the cases in the data, but which could be dismissed as common-sense, routine decisions and not in need of any pretentious “analysis”), their model still predicts almost 83% of reported behavior. Even for the second stage, after first-stage decisions fail to result in a cure, and decisions get more complex and tougher to predict, the model predicts an impressive 84% of reported behavior.

Ryan and Martínez's EDM: IF-THEN Charts

Gery Ryan and Homero Martínez (1996) built an EDM for how mothers in San José, Mexico, treated children who have diarrhea. Ryan and Martínez knew, from living in the village, that mothers there use seven different treatments in treating their children's diarrhea. Five of the treatments consist of giving the child one or more of the following: (1) tea, (2) homemade rice water, (3) medication from the pharmacy (their informants told them "If you can say it, you can buy it"), (4) a carbonated beverage, or (5) a commercially produced oral rehydration solution. The other two treatments are: (6) manipulating the child's body (massaging the child's body, pinching the child's back) or (7) taking the child to the doctor.

Ryan and Martínez asked 17 mothers in San José who had children under age 5 what they did the last time their children had diarrhea. Then they went systematically through the treatments, asking each mother why she had used X instead of A, X instead of B, X instead of C, and so on down through the list.

Mothers in San José listed the following factors for choosing one treatment over another:

- duration of the episode
- perceived cause (from worms, from *empacho*, from food, etc.)
- whether there was mucous in the stool
- whether there was blood in the stool
- whether the stools smelled bad
- whether the stools were frequent or not
- whether the stools were loose or not
- whether the child had fever
- color of the stool
- whether the child had a dry mouth
- whether the child had dry eyes
- whether the child was vomiting
- whether the child had swollen glands

Table 18.4 shows the data from the 17 women in Ryan and Martínez's original sample and the decision to take the child to the doctor. Read the table like this: Mother #1 said that her child's last episode of diarrhea lasted 2 days and was caused by bad food. The stools contained mucous, but did not contain blood. The stools smelled bad, were frequent and loose. The child had fever, the stools were yellow. The child had dry mouth and dry eyes, but was not vomiting and did not have swollen glands. In the end, Mother #1 did not take

TABLE 18.4
Decisions to Take a Child to the Doctor in San José, Mexico

<i>mother</i>	<i>doctor</i>	<i>days</i>	<i>cause</i>	<i>muc.</i>	<i>blood</i>	<i>smell</i>	<i>freq</i>	<i>loose</i>	<i>fever</i>	<i>color</i>	<i>mouth</i>	<i>eyes</i>	<i>vomit</i>	<i>gland</i>
1	N	2	C	Y	N	Y	Y	Y	Y	A	Y	Y	N	N
2	N	20	E	Y	N	Y	Y	Y	N	A	Y	Y	N	N
3	Y	8	T	N	N	Y	Y	Y	Y	N	Y	Y	N	N
4	Y	8	C	Y	N	Y	N	Y	Y	V	.	.	Y	Y
5	N	3	P	Y	N	Y	Y	Y	Y	A	Y	Y	N	Y
6	N	3	L	N	N	Y	Y	Y	N	B	Y	Y	N	N
7	Y	8	D	Y	N	Y	Y	Y	N	A	Y	Y	N	N
8	N	1	D	N	N	Y	Y	Y	N	A	.	.	N	N
9	N	.	C	Y	N	N	Y	Y	N	B	Y	Y	N	N
10	N	3	O	N	N	Y	Y	Y	N	A	Y	Y	N	N
11	N	2	C	N	N	N	N	N	N	A	.	.	N	N
12	N	.	C	N	N	Y	Y	Y	N	A	Y	Y	N	N
13	N	4	C	N	N	Y	N	N	N	A	Y	Y	N	N
14	Y	4	E	N	N	Y	Y	Y	Y	V	Y	.	N	N
15	Y	3	I	Y	Y	Y	Y	Y	Y	A	Y	Y	Y	N
16	N	2	C	Y	N	N	Y	Y	N	V	Y	Y	N	N
17	N	7	E	N	N	N	Y	Y	N	A	N	N	N	N

Cause		Color
C = food	D = teething	A = yellow
L = worms	T = dirt	V = green
E = <i>empacho</i>	P = parasites	B = white
I = indigestion	O = other	N = black

SOURCE: G. W. Ryan and H. Martínez (1996). Data used with permission of the authors.

her child to the doctor. (The codes for cause and color are from Spanish; see the legend just below the table.)

Table 18.4 makes it clear that mothers took their children to the doctor if the child had blood in the stool, had swollen glands, or was vomiting, or if the diarrhea had lasted *more than 7* days. None of the other factors played a part in the final decision to take the child to the doctor.

But remember: There were seven different treatments, and mothers often try several treatments in any given episode. Ryan and Martínez looked at the pattern of circumstances for all seven treatments and built a model that accounted for the treatment decisions made by the 17 mothers. Their model had just six rules and three constraints.

Figure 18.4 shows the model as a series of IF-THEN statements. Notice the constraints: For a woman to choose a modern medication, she has to know about it and it has to be easy to get and cheap. The constraints to the rules are derived from ethnographic interviews. So was the observation that mothers distinguished between curative and palliative treatments—treatments that stop diarrhea and treatments that simply make the child feel better until the episode

	<u>Rule 1</u>
IF	Child has blood stools OR child has swollen glands OR child is vomiting
THEN	take child to doctor.
	<u>Rule 2</u>
IF	diarrhea is caused by <i>empacho</i>
THEN	give physical treatment.
	<u>Rule 3</u>
IF	previous rules do not apply OR there is no cure with <i>empacho</i> treatment
THEN	give the highest preferred curing treatment that meets constraints.
	<u>Rule 4</u>
IF	previous treatment did not stop diarrhea
THEN	compare the two highest treatments of remaining options.
	<u>Rule 4.1</u>
IF	one is a curing remedy AND meets its constraints
THEN	give this treatment.
	<u>Rule 4.2</u>
IF	both or neither are curing remedies AND each meets its respective constraints
THEN	give the highest ranked preference.
	<u>Rule 5</u>
IF	the previous treatment did not stop the diarrhea AND
THEN	the episode is less than 1 week repeat rule 4
	<u>Rule 6</u>
IF	the episode has lasted more than 1 week
THEN	take the child to a doctor.
	<u>Constraints</u>
IF	you know how to make ORS (oral rehydration solution) AND
THEN	your child will drink ORS give ORS
IF	you know a medication that works for diarrhea AND
THEN	you have it in the house give the pill or liquid medication.
IF	you know a medication that works for diarrhea AND
THEN	it is cheap AND it is easy to obtain. give the pill or liquid medication.

Figure 18.4. Ryan and Martínez's decision model as a series of IF-THEN rules.

SOURCE: G. W. Ryan and H. Martínez, "Can We Predict What Mothers Do? Modeling Childhood Diarrhea in Rural Mexico," *Human Organization*, Vol. 55, pp. 47-57, 1996. Reprinted with permission of the Society for Applied Anthropology.

is over. The model **postdicted** (accounted for) 89% of the treatments that the 17 mothers had reported.

Next, Ryan and Martínez tested their model. They interviewed 20 more mothers, but this time they asked each woman every question in the model. In other words, they asked each woman: “In your child’s last episode of diarrhea, did the stools have blood in them? Did the child have swollen glands? What caused the diarrhea?” and so on. The IF-THEN model in figure 18.4 accounted for 84% of the *second* group’s treatment decisions.

Folk Taxonomies

There are about 6,000 languages spoken in the world today. Speakers of all those languages name things in the natural world. In 1914, Henderson and Harrington published a monograph on the ethnozoology of the Tewa Indians of New Mexico. Scholars ever since have been interested in understanding the variety of ways in which people organize their knowledge of the natural world.

In the 1950s, anthropologists began systematically producing folk taxonomies—that is, hierarchical, taxonomic graphs to represent how people organize their knowledge of plants and animals. These ethnobotanical and ethnozoological taxonomies don’t necessarily mirror scientific taxonomies, but then, the whole point of what became known as **ethnoscience** is to understand cultural knowledge on its own terms.

Scientific taxonomies for plants and animals recognize six primary levels of distinction (phylum, class, order, family, genus, and species) and lots of in-between levels, as well (infraorder, superorder, subclass, etc., etc.), but folk taxonomies of plants and animals across the world are generally limited to five or, at most, six levels. Figure 18.5 (from D’Andrade 1995) shows part of the folk taxonomy of *creatures* for native speakers of English.

Covert Categories

There are six culturally appropriate levels of hierarchical distinction identified in figure 18.5: (1) First, there is the unique beginner, a single label that identifies the cultural domain. (2) Next comes a relatively small set of life forms (animals, fish, insects, etc.). (3) Then there is an intermediate level, which includes **covert categories**, if any exist in a particular taxonomy. Folk genera (level 4), folk species (level 5), and folk varieties (level 6) round out the picture.

There is a covert, unnamed category in figure 18.5 comprising wolves, foxes, dogs, coyotes, and some other things (the dashed line extending down

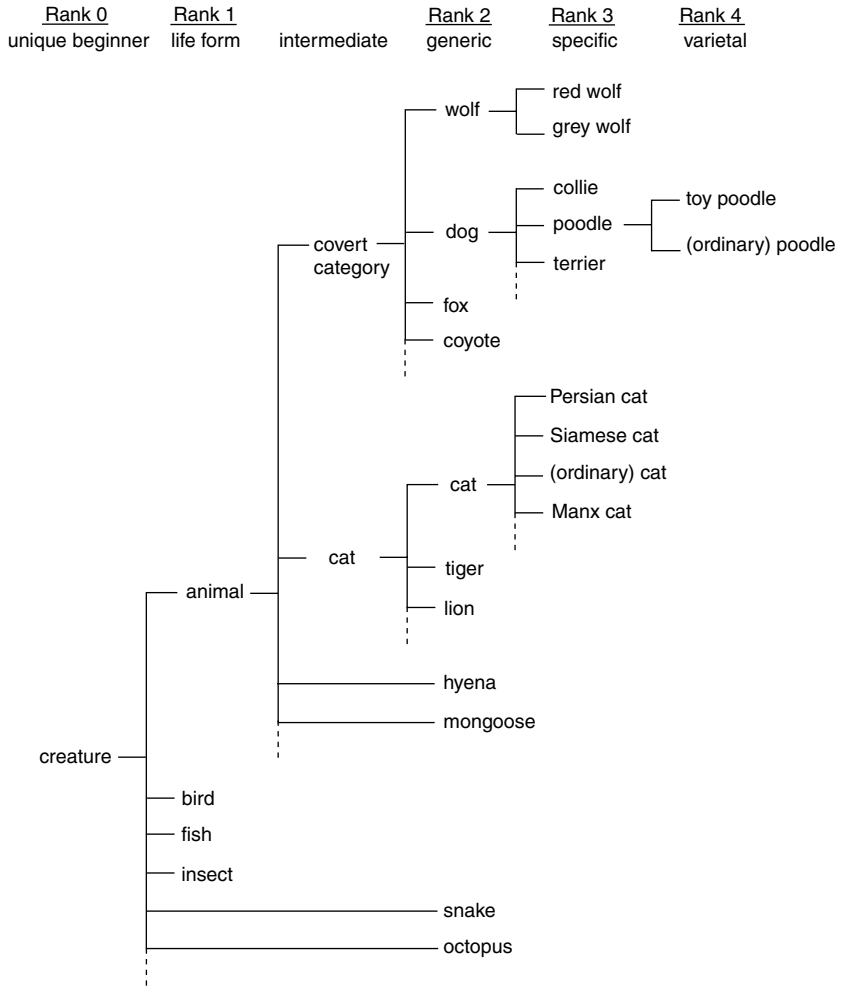


Figure 18.5. Partial taxonomy for *creatures* in English.

SOURCE: R. G. D'Andrade, *The Development of Cognitive Anthropology*, p. 99. © Cambridge University Press 1995. Reprinted with the permission of Cambridge University Press.

from coyote, indicates that the covert category contains more than what's listed in the figure). In a scientific taxonomy, foxes are not in the same genus with dogs and wolves. The latter are in the genus *Canis*, while foxes are in the genus *Vulpes*. Many speakers of English, however, classify foxes and wolves in the category of “things in the dog family,” or “canines,” and a folk taxonomy of English animal terms respects that.

The intermediate category of “cat” is not covert. How can you tell? As D'Andrade says, you can say “Look at that cat!” if you're talking about a tiger, but it's weird to say “Look at that dog” if you're pointing to a fox, so “cat” is a named intermediate category and “dog” isn't.

Two more things about figure 18.5. Note how we use words for generic animals in English that would be at the species level in a scientific taxonomy (wolf, coyote, and dog are all members of the genus *Canis*, species *lupus*, *latrans*, and *familiaris*, respectively), and how the species level in the folk taxonomy comprises names for subspecies in a scientific taxonomy.

Finally, look at how D'Andrade has placed octopus and snake in figure 18.5. The horizontal lines show that D'Andrade has classified these creatures as **nonaffiliated generics**. They might be classified as life forms, but, as D'Andrade points out, there are many nonaffiliated generics in the ocean, including clams, lobsters, seahorses, jellyfish, and octopi.

Cultural Domains and Folk Taxonomies

It was quickly recognized that folk taxonomies could be developed for *any cultural domain*, not just for ethnobotanical and ethnozoological knowledge, and that we use folk taxonomies all the time to order our experience and guide our behavior.

Take someone to a supermarket—one they've never been to before—and ask them to find peanut butter. Follow them as they make their way around the store and get them to talk about what they think they're doing. Here's a typical response:

Well, let's see, milk and eggs are over there by that wall, and the meat's usually next to that, and the canned goods are kind of in the middle, with the soaps and paper towels and stuff on the other side, so we'll go right in here, in the middle. No, this is the soap aisle, so let's go over to the right. Sure, here's the coffee, so it's got to be on this aisle or the next, with cans of things like ravioli.

Any competent member of U.S. or Canadian culture will find the peanut butter in a hurry, but not everything is so clear. Shredded coconut and walnuts are often shelved with flour in the United States because they are used in bak-

ing, while other nuts—cashews and peanuts, for example—may be shelved somewhere else, like with the snacks. Lychee nuts (a Chinese dessert food) and matzohs (unleavened bread boards eaten primarily by Jews) are sometimes shelved in American supermarkets together under “ethnic foods,” but may be shelved in separate “Oriental foods” and “Jewish foods” sections if local populations of those groups are sufficiently large.

Spradley (1979) reported that he once called the St. Paul, Minnesota, police department and said he needed to find the case number of a robbery that had been committed at his house. Two bicycles had been stolen from his garage in the middle of the night, while he was asleep. The police had investigated, but Spradley’s insurance company needed the case number to process the claim. When Spradley told the police that he needed the case number for a “robbery,” they quite naturally transferred his call to the robbery unit. But the people there couldn’t help him because, according to their rules, robberies involve a face-to-face encounter between the criminal and the victim and the criminal uses a gun.

Spradley was transferred to burglary, but they couldn’t help him either because, they said, theft of bicycles is handled by the juvenile division in St. Paul. Eventually, Spradley got his case number, but, he said, if he had understood the police culture, he “would have begun with a simple question: What part of the police department has records of bicycles stolen from a garage when no one is present?” (1979:142). In other words, if he’d known the taxonomy for the cultural domain of crimes, he’d have asked the right question and gotten taken care of right away.

Cultural domains are everywhere. Ask people to free list any set of things you can name—emotions, names of ethnic groups, junk foods, hand tools. If lots of people can respond easily, you’ve gotten hold of a cultural domain. Some domains (like animals and plants) are huge; others (like the names of carpenters’ tools) are small but require specialized knowledge.

Some domains are highly consensual. All native speakers of Spanish know the list of names of family members (*madre, padre, hijo, hija*, etc.). There are some specialized uses of terms that vary from one Spanish-speaking country to another, but no student of any Spanish-speaking culture could avoid learning that domain. (See Borgatti [1993/1994] for a seminal discussion of cultural domain analysis. For classic work on folk taxonomies, see the many contributions in Frake 1961, Tyler 1969, and Werner and Schoepfle 1987.)

How to Make a Taxonomy: Lists and Frames

In building a folk taxonomy, many researchers combine the free-list technique I described in chapter 11 and the **frame elicitation** technique developed

by Frake (1964), Metzger and Williams (1966), and D'Andrade et al. (1972). Start with the frame:

What kinds of _____ are there?

where the blank is “cars,” “trees,” “saddles,” “snow,” “soldiers”—whatever you’re interested in understanding. This frame is used again and again, until an informant says that the question is silly.

For example, suppose you asked a native speaker of American English “What kinds of foods are there?” You might get a list like: pasta, meat, fish, fruits, vegetables, snacks. . . .” (You’ll probably get a slightly different set of labels if you ask a native speaker of British English this same question.)

Next, you ask: “What kinds of pasta [meats] [fish] [etc.] are there?” The answer for meats might be: beef, lamb, chicken, pork, venison. . . .

So you extend the search: “What kinds of beef [lamb] [chicken] [etc.] are there?” For some people, at least, you’ll find that beef is divided into steak, chops, hamburger, and so on, and that chicken is divided into dark meat and white meat. But if you ask “What kinds of steaks are there?” you might be told: “There are no kinds; they just are what they are.” Of course, if you’re dealing with a real steak lover, you might be told about Porterhouse, T-bone, rib eye, Delmonico, filet mignon, and so on.

Once you have a list of lexical items in a domain, and once you’ve got the basic divisions down, the next step is to find out about overlaps. Some foods, like peanuts, get classified as snacks and as protein sources by different people—or even by the same person at different times.

The point is, while the Food and Drug Administration may have codified foods in the United States, there is no codified set of folk rules for a taxonomy of foods in American culture. The only way to map this is to construct folk taxonomies from information provided by a number of people and to get an idea of the range of variation and areas of consistency in how people think about this domain. You can learn about the possible overlaps in folk categories by using the substitution frames:

Is _____ a kind of _____?

Is _____ a part of _____?

Once you have a list of terms in a domain, and a list of categories, you can use this substitution frame for all possible combinations. Are marshmallows a kind of meat? A kind of fish? A kind of snack? This can get really tedious, but discovering **levels of contrast**—that *magenta* is a kind of *red*, that *cashews* are a kind of *nut*, that *alto* is a kind of *sax*, or that *ice cream* is a kind

of *dessert*—just takes plain hard work. Unless you're a child, in which case all this discovery is just plain fun.

A common way to display folk taxonomies is with a branching tree diagram. Figure 18.6 shows a tree diagram for part of a folk taxonomy of passenger cars. I elicited this taxonomy in Morgantown, West Virginia, from Jack in 1976.

Things to Look for in Folk Taxonomies

There are four important points to make about the taxonomy shown in figure 18.6:

1. **Interinformant variation** is common in folk taxonomies. That is, different people may use different words to refer to the same category of things. Sometimes, in fact, terms can be almost idiosyncratic. Jack distinguished among what he called “regular cars,” “station wagons,” and “vans.” The term “regular cars” is not one you normally see in automobile ads, or hear from a salesperson on a car lot.
2. Category labels do not necessarily have to be simple lexical items, but may be complex phrases. The category labeled “4-wheel drive” vehicles in figure 18.6 was sometimes called “off-road vehicles” in 1976, or even “vehicles you can go camping in or tow a horse trailer with.” Jack said that Jeep station wagons were both wagons *and* 4-wheel-drive cars you can go camping in. Remember, I elicited this taxonomy in 1976, long before the advertising industry came up with the concept of a “sport utility vehicle.” In 1998, I asked some people to list “types of cars.” The term “utes” showed up and I flat didn't understand it. Of course, after expressing my ignorance, I learned that “utes” is short for “sport utility vehicles,” better known as SUVs.
3. There are those covert categories I mentioned—categories for which people have no label at all, or at least not one they find easily accessible. Some people insist that Corvettes, Camaros, Maseratis, and MGs are part of a single category, which they find difficult to name (one informant suggested “sporty cars” as a label). Others, like Jack, separate “performance cars” from “sports cars” and even subdivide sports cars into “true sports cars” and “rally cars.” Be on the lookout for unlabeled categories (that is, unlabeled nodes in a branching tree diagram) in any folk taxonomy.
4. Even when there are consistent labels for categories, the categories may represent multiple dimensions, each of which has its own levels of contrast. For example, many native speakers of American English recognize a category of “foreign cars” that cuts across the taxonomy in figure 18.6. There are foreign sports cars, foreign luxury cars, and foreign regular cars.

Folk taxonomies can be very, very complex. One way to get at the complexity is through a technique known as componential analysis. (Another is multi-dimensional scaling, which we'll get to in chapter 21.)

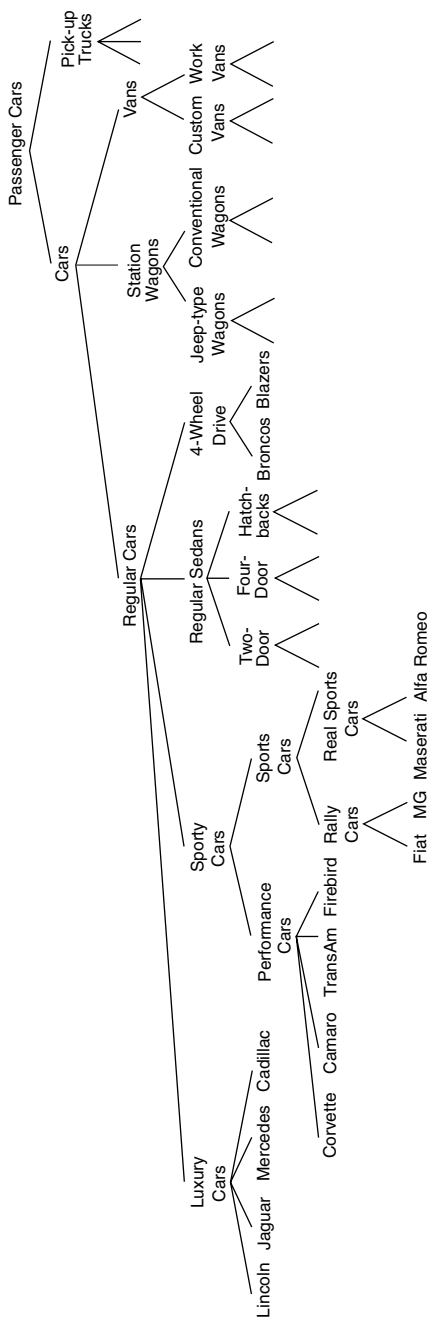


Figure 18.6. Part of Jack's taxonomy of cars and trucks.

Componential Analysis

Componential analysis is a formal, qualitative technique for studying meaning. There are two objectives: (1) to specify the conditions under which a native speaker of a language will call something (like a plant, a kinsman, a car) by a particular term and (2) to understand the cognitive process by which native speakers decide which of several possible terms they should apply to a particular thing.

The first objective is descriptive, but the second is a kind of **causal analysis** and is what the developers of the technique had in mind in the 1950s and 1960s (see Conklin 1955; Goodenough 1956; Frake 1962; Wallace 1962). Charles Frake, for example, described componential analysis as a step toward “the analysis of terminological systems in a way which reveals the conceptual principles that generate them” (1962:74).

Componential analysis is based on the principle of **distinctive features** in phonology, the branch of linguistics devoted to the study of the sounds of a language. To understand the principle, think about the difference in the sounds represented by P and B in English. Both are made by twisting your mouth into the same shape. This is a *feature* of the P and B sounds called “bilabial” or “two-lipped.”

Another feature is that they are both “stops.” That is, they are made by stopping the flow of air for an instant as it moves up from your lungs and releasing the flow suddenly. An S sound, by contrast, also requires that you restrict the air flow, but not completely. You kind of let the air slip by in a hiss. The only difference between a P and a B sound is that the P is voiceless while the B is voiced—you vibrate your vocal cords while making a B.

If you add up all the phonological features of the words “bit” and “pit,” the only feature that differentiates them is voicing on the first sound in each word. The “pitness” of a pit and the “bitness” of a bit are clearly not in the voicelessness or voicedness of the sounds P and B; but any native speaker of English will distinguish the two words, and their meanings, and can trace the difference between them to that little feature of voicing if you push them a bit.

There is a unique little bundle of features that define each of the consonantal sounds in English. The only difference between the words “mad” and “bad” is that the bilabial sound M is nasal, and not a stop. These distinctive features carry meaning for native speakers of a language.

This principle can be adapted to the study of other domains of culture. Any two “things” (sounds, kinship terms, names of plants, names of animals, etc.) can be distinguished by exactly one binary feature that either occurs (+) or doesn’t occur (–). Table 18.5 shows that with two features you can distinguish four things: Thing 1 can be ++, thing 2 can be + –, thing 3 can be

TABLE 18.5
A Componential Analysis of Four Things with Two Features

	<i>Feature 1</i>	<i>Feature 2</i>
Thing 1	+	+
Thing 2	+	–
Thing 3	–	+
Thing 4	–	–

– +, and thing 4 can be – –. Each bundle of features is different and defines each of the four things. With three binary features, you can distinguish 8 things; with four, 16; with five, 32; and so on.

When componential analysis was introduced into cultural anthropology, it was applied to the set of English kinship terms (Goodenough 1956) and it continues to be used for understanding kinship systems (Rushforth 1982; Kendall 1983; Hedican 1986; Pericliev and Valdez-Perez 1998). A “daughter” in English, for example, is a *consanguineal, female, descending generation* person. So is a niece, but a niece is *through a sibling or a spouse*.

Table 18.6 shows the distinctive feature principle applied to a list of 22 barnyard animals. Stallions are adult male horses and foals are baby horses. Notice that there is no column in table 18.6 labeled “male” and no column labeled “juvenile.” A parsimonious set of features for distinguishing among these animals does not require all that information. A stallion is a nonfemale, gendered (that is, not neutered), adult horse. Any horse that is not female and gendered and not an adult must be a colt. A barrow is a neutered, adult hog, and a wether is a neutered, adult sheep.

Actually, if we wanted the most parsimonious set of features we would drop one of the last four columns in table 18.6. Those columns identify the general class of animals—horses, cattle, sheep, and swine—to which each named animal belongs. If the domain of barnyard animals comprised *just* those four classes, then a not-horse, not-cattle, not-sheep animal *must* be a swine. I’ve left all four animal classes in table 18.6 because there are other classes of barnyard animals not yet represented (chickens, rabbits, goats, etc.).

Componential analysis can be applied to any domain of a language where you are interested in understanding the semantic features that make up the domain. Table 18.7 shows a componential analysis of seven cars, using three features elicited from Jack (he of the taxonomy shown in figure 18.6). A Corvette is an expensive car, not very practical, and not foreign; a Mercedes is an expensive, practical, foreign car; and so on. Each of the seven cars is uniquely defined by the three features Jack mentioned.

TABLE 18.6
A Componential Analysis of 17 Barnyard Animals

	<i>Female</i>	<i>Neuter</i>	<i>Adult</i>	<i>Horses</i>	<i>Cattle</i>	<i>Sheep</i>	<i>Swine</i>
Cow	+	-	+	-	+	-	-
Bull	-	-	+	-	+	-	-
Steer	-	+	+	-	+	-	-
Calf	-	+	-	-	+	-	-
Heifer	+	-	-	-	+	-	-
Mare	+	-	+	+	-	-	-
Stallion	-	-	+	+	-	-	-
Gelding	-	+	+	+	-	-	-
Foal	-	+	-	+	-	-	-
Filly	+	-	-	+	-	-	-
Colt	-	-	-	+	-	-	-
Sow	+	-	+	-	-	-	+
Boar	-	-	+	-	-	-	+
Barrow	-	+	+	-	-	-	+
Piglet	-	+	-	-	-	-	+
Gilt	+	-	-	-	-	-	+
Shoat	-	-	-	-	-	-	+
Sheep	-	-	+	-	-	+	-
Ram	-	-	+	-	-	+	-
Ewe	+	-	+	-	-	+	-
Wether	-	+	+	-	-	+	-
Lamb	-	-	-	-	-	+	-

TABLE 18.7
Minimal Componential Analysis for Seven Cars, According to Jack

	<i>Expensive</i>	<i>Practical</i>	<i>Foreign</i>
Corvette	+	-	-
Firebird	-	-	-
MG	-	-	+
Maserati	+	-	+
Mercedes	+	+	+
Jeep	-	+	-
Dodge Van	+	+	-

There are two problems with componential analysis. First of all, it seems a bit shallow to say that a Corvette is an expensive, impractical, American car and nothing more, or that a Mercedes is an expensive, practical, foreign car and nothing more. You can get so caught up in finding the minimal analytic combination of features in this type of analysis that you forget you're interested in the meaning that people assign to different objects in a domain. On the other hand, if you know the most parsimonious set of distinctive features for an item in a domain, you can predict how someone will label new things in the domain that they haven't encountered before.

The second problem with componential analysis is the same one we run into with all cognitive research methods: We have no idea if it reflects how people actually think. This problem was raised early in the development of cognitive studies by Robbins Burling (1964), who noted that, in a folk taxonomy of trees, he could not tell the essential cognitive difference between hemlock and spruce. "Is it gross size, type of needle, form of bark, or what?" If an ethnographer could not answer this question, Burling observed, then no componential analysis could claim to be "more than an exercise of the analyst's imagination" (p. 27).

Of course, this same critique could apply to any social research that "imputes the presence of something inside people" (like values and attitudes) and must be balanced with a positive perspective on what *can* be done (Hymes 1964:119).

In fact, what can be done is impressive, intuitively compelling analysis of the meanings that people attach to terms in their languages: Decision analysis allows us to predict which of several behavioral options people will take, under specific circumstances; taxonomic analysis lets us predict which class of things some new thing will be assigned to; componential analysis lets us predict what classification label will be assigned to some object. These methods produce effective knowledge that finds application in practical fields like health care delivery and advertising.

Analytic Induction and Boolean Tests

Analytic induction is a formal, qualitative method for building up causal explanations of phenomena from a close examination of cases. The method involves the following steps: (1) Define a phenomenon that requires explanation and propose an explanation. (2) Examine a single case to see if the explanation fits. (3) If it does, then examine another case. An explanation is accepted until a new case falsifies it.

When you find a case that doesn't fit, then, under the rules of analytic induc-

tion, the alternatives are to change the explanation to include the new case or redefine the phenomenon to exclude the nuisance case. Ideally, the process continues until a universal explanation for all known cases of a phenomenon is attained. (Explaining cases by declaring them all unique is not an option of the method. That's a convenient way out, but it doesn't get us anywhere.)

Charles Ragin (1987, 1994) formalized the logic of analytic induction, using an approach based on **Boolean logic**. Boolean variables are dichotomous: true or false, present or absent, and so on. This seems simple enough, but it's going to get very complicated, very quickly, so pay attention. Remember, there is no math in this. It's entirely qualitative. In fact, Ragin (1994) calls his Boolean method of induction **qualitative comparative analysis**, or QCA.

Suppose you have four dichotomous variables, including three independent, or causal, variables, and one independent, or outcome, variable. With one dichotomous variable, A, there are 2 possibilities: A and not-A. With two dichotomous variables, A and B, there are 4 possibilities: A and B, A and not-B, not-A and B, not-A and not-B. With three dichotomous variables, there are 8 possibilities; with four there are 16 . . . and so on.

We've seen all this before: in the discussion about factorial designs of experiments (chapter 5); in the discussion of how to use the number of subgroups to figure out sample size (chapter 6); in the discussion of how to determine the number of focus groups you need in any particular study (chapter 9); and in the discussion of factorial questionnaires (chapter 10). The same principle is involved.

Thomas Schweizer (1991, 1996) applied this Boolean logic in his analysis of conflict and social status in Chen Village, China. In the 1950s, the village began to prosper with the application of technology to agriculture. The Great Leap Forward and the Cultural Revolution of the 1960s, however, reversed the village's fortunes. Chan et al. (1984) reconstructed the recent history of Chen Village, focusing on the political fortunes of key actors there.

Schweizer coded the Chan et al. text for whether each of 13 people in the village experienced an increase or a decrease in status after each of 14 events (such as the Great Leap Forward, land reform and collectivization, the collapse of Red Brigade leadership, and an event known locally as "the great betrothal dispute"). Schweizer wound up with a 13-actor-by-14-event matrix, where a 1 in a cell meant that an actor had success in a particular event and a 0 meant a loss of status in the village.

When Schweizer looked at this actor-by-event matrix he found that, over time, nine of the actors consistently won or consistently lost. That means that for nine of the villagers, there was just one outcome, a win or a loss. But four of the actors lost *sometimes* and won other times. For each of these four peo-

ple, there could be a win or a loss, which means there are eight possible outcomes.

In total, then, Schweizer needed to account for 17 unique combinations of actors and outcomes. He partitioned the 17 unique cases according to three binary independent variables (whether a villager was originally from a city or had been raised in the village, whether a villager had a proletarian or a nonproletarian background, and whether a villager had ties to people outside the village or not) and one dependent variable (whether the person was an overall success). There are a total of four variables. Table 18.8 shows the 16

TABLE 18.8
The Outcome of 17 Cases from Schweizer's (1996) Text Analysis

<i>Success</i>	<i>External ties</i>	<i>Proletarian background</i>	<i>Urban origin</i>	<i>No. of cases</i>
0	0	0	0	2
0	0	0	1	2
0	0	1	0	1
0	0	1	1	0
0	1	0	0	0
0	1	0	1	0
0	1	1	0	2
0	1	1	1	0
1	0	0	0	1
1	0	0	1	3
1	0	1	0	0
1	0	1	1	0
1	1	0	0	0
1	1	0	1	1
1	1	1	0	4
1	1	1	1	1

outcomes that are possible with four binary variables, and the number of actual cases, out of 17, for each of those outcomes.

By setting up the logical possibilities in table 18.8, Schweizer was able to test several hypotheses about success and failure in Chen Village. You can see, for example, that just two people who were originally from a city (people who had been sent to the village to work during the Cultural Revolution) turned out to be failures, but five people from the city turned out to be successful. People from an urban background have an advantage, but you can also see from table 18.8 that it's not enough. To ensure success, you should come from a proletarian family OR have good external ties (which provide access to information and power at the regional level).

Failure is predicted even better: If an actor has failed in the Chen Village disputes, then he or she is of rural origin (comes from the village) OR comes from a nonproletarian family AND has no ties to authorities beyond the village. The Boolean formula for this statement is:

Lack of success → *v nonurban (nonproletarian and lack of ties)*

The substantive conclusions from this analysis are intuitively appealing: In a communist revolutionary environment, it pays over the years to have friends in high places; people from urban areas are more likely to have those ties; and it helps to have been born into a politically correct (that is, proletarian) family.

Analytic induction helps identify the simplest model that logically explains the data. Like classic content analysis and cognitive mapping, human coders have to read and code the text into an event-by-variable matrix. The object of the analysis, however, is not to show the relations between all codes, but to *find the minimal set of logical relations* among the concepts that accounts for a single dependent variable.

With three binary independent variables (as in Schweizer's data), two **logical operators** (OR and AND), and three implications ("if A then B," "if B then A," and "if A, then and only then, B"), there are 30 multivariate hypotheses: 18 when all three independent variables are used, plus 12 when two variables are used. With more variables, the analysis becomes much more difficult, but there are now computer programs, like Tosmana and *Anthropac* (Borgatti 1992a, 1992b), that test all possible multivariate hypotheses and find the optimal solution (see appendix F).

Here are the details of the Boolean logic of Schweizer's analysis (Schweizer 1996). Three possible hypotheses can be derived from two binary variables: "If A then B," "If B then A," and "If A, then and only then, B." In the first hypothesis, A is a sufficient condition to B and B is necessary to A. This hypothesis is falsified by all cases having A and not B. In the second hypothesis, B is a sufficient condition to A and A is necessary to B. The second hypothesis is falsified by all cases of B and not A. These two hypotheses are **implications** or **conditional statements**. The third hypothesis (an **equivalence** or **biconditional statement**) is the strongest: Whenever you see A, you also see B and vice versa; the absence of A implies the absence of B and vice versa. This hypothesis is falsified by all cases of A and not B, and all cases of B and not A.

Applied to the data from Chen Village, the strong hypothesis is falsified by many cases, but the sufficient condition hypotheses (urban origin implies success; proletarian background implies success; having external ties implies success) are true in 86% of the cases (this is an average of the three sufficient condition hypotheses). The necessary condition hypotheses (success implies

urban origin; success implies proletarian background; success implies external ties) are true in just 73% of cases (again, an average). (There are seven disconfirming cases in 51 possible outcomes of the 12 **sufficient condition** possibilities—four possible outcomes for each of three independent variables and one dependent variable. There are 14 disconfirming cases in 51 possible outcomes of the 12 **necessary condition** possibilities.) To improve on this, Schweizer tested multivariate hypotheses, using the logical operators OR and AND (Schweizer 1996).

These are not easy analyses to do, and some people I talk to about this kind of work wonder how qualitative analysis ever got so complicated. It just goes to show that qualitative doesn't mean wimpy. (For more examples of QCA, see Ragin 1998, Haworth-Hoepfner 2000, Rantala and Hellstrom 2001, Dixon et al. 2004, Hodson 2004, and Roscigno and Hodson 2004.)

19



Univariate Analysis

The next three chapters deal with methods for analyzing quantitative data. We begin with descriptive and inferential univariate analysis. Then, in chapters 20 and 21, we move on to bivariate and multivariate analysis.

Descriptive analysis involves understanding data through graphic displays, through tables, and through summary statistics. Descriptive analysis is about the data you have in hand. **Inferential analysis** involves making inferences about the world beyond the data you have in hand.

When you say that the average age of people in a village is 44.6 years, that's a descriptive analytic statement. When you say that there is a 95% probability that the true mean of the population from which you drew a sample of people is between 42.5 and 47.5 years, that's an inferential statement: You are inferring something about a population from data in a sample.

In **univariate analysis**, we examine variables precisely and in detail and get to know the data intimately. **Bivariate analysis** involves looking at associations between pairs of variables and trying to understand how those associations work. **Multivariate analysis** involves, among other things, understanding the effects of more than one independent variable at a time on a dependent variable.

Suppose you're interested in the causes of variation in the income of women. You measure income as the dependent variable and some independent variables like: age, marital status, employment history, number of children, ages of children, education, and so on. The first thing to do is examine carefully the properties of all the variables. That's the univariate part of the analysis.

Next, you'd look at the association between each independent variable and the dependent variable. You'd also look at the association between pairs of

independent variables. That's the bivariate part. Finally, you'd look at the simultaneous effect of the independent variables on the dependent variable or variables. That's the multivariate part.

Each part helps us answer questions about how things work.

Coding and Codebooks for Quantitative Data

Quantitative data processing depends crucially on having a useful **codebook**. A codebook for quantitative data spells out exactly how to transform observations into numbers that can be manipulated statistically and searched for patterns.

A good codebook is worth a lot in data analysis and it's worth more every year. It tells you (and others) what you have in your data—what variables you've studied, what you've called those variables, and how you've stored information about them. You simply can't analyze quantitative data without a good, clear codebook.

Just as important, neither can anyone else. You can't share your data with other researchers unless you give them a codebook they can use. Six months after you finish anything but the simplest projects (those with only half a dozen or fewer variables), even *you* won't recognize your own data without a codebook. And if you want to reanalyze your data several years after a project has ended, or compare data from 2002 with current data, you won't be able to do so unless you have built and filed away a good codebook.

Coding

The first rule for coding quantitative data is: Don't analyze while you're coding. This rule is the exact opposite of the rule that applies to coding qualitative data. Coding text *is* analysis—thinking about what each piece of text means, developing hypotheses about the people who are described, boiling the text down to a series of mnemonics.

It's different with a set of numbers. Suppose you ask 400 randomly selected people, aged 20–70, how old they are. You could get as many as 51 different ages, and you'll probably get at least 20 different ages.

I've seen many researchers code this kind of data into four or five categories—such as 20–29, 30–39, 40–49, 50 and older—before seeing what they've got. Recall from chapter 2 that this just throws away the interval-level power of data about age. You can always tell the computer to package data about age (or income, or any interval variable) into a set of ordinal chunks.

But if you actually code the data into ordinal chunks to begin with, you can never go back.

Here's a concrete example of something that's a little more complex than age. Gene Shelley studied the strength of ties between friends and acquaintances (Shelley et al. 1990). Every other day for a month, she called 20 informants on the phone to talk about things they'd learned in the previous two days about their friends and acquaintances. People mentioned things like "So-and-so told me she was pregnant," "So-and-so's father called and told me my friend made his first jump in parachute school," and so on. Shelley asked people to estimate how long it had been between the time something happened to one of their friends/acquaintances and the time they (the informants) heard about it. This estimated time was the major dependent variable in the research.

There were 20 informants, who submitted to 15 interviews each, and in each interview almost every informant was able to name several events of interest. Thus, there were over 1,000 data records (one for each event remembered by an informant). The length of time estimated by informants between an event happening to someone they knew and their hearing about it ranged from "immediately," to "10 years," with dozens of different time periods in between ("about 5 minutes," "two and a half months," etc.).

The temptation was to make up about five codes, like 1 = 5 minutes or less, 2 = 6 minutes to 19 minutes, 3 = 20 minutes to an hour, and so on. But how do you decide what the right breaks are? Shelley decided to code everything in *days* or *fractions of days* (1 minute is .0007 days; 10 years is 3,650 days) without worrying about leap years (ibid.). Shelley didn't throw away data by turning a ratio-level variable (minutes) into an ordinal variable (arbitrary chunks of time).

Here's another example, using a nominal variable. Suppose you are studying the personal histories of 200 Mexican men who have had experience as illegal labor migrants to the United States. If you ask them to name the towns in which they have worked, you might get a list of 300 communities—100 more than you have informants! The temptation would be to collapse the list of 300 communities into a shorter list, using some kind of scheme. You might code them as Southeast, Southwest, California, Midwest, Northwest, mid-Atlantic, and so on. Once again, you'd be making the error of doing your analysis in the coding.

Once you've got all the data entered into a computer, you can print them, lay them out, stare at them, and start making some decisions about how to "package" them for statistical analysis. You might decide to label each of the 300 communities in the list according to its population size; or according to its ethnic and racial composition (more than 20% Spanish surname, for example); or its distance in kilometers from the Mexican-U.S. border. All those

pieces of information are available from the U.S. Census or from road atlases. But if you collapse the list into a set of categories during coding, then your option to add codes about the communities is closed off.

Building Codebooks

Table 19.1 shows the codebook for a street-intercept survey that some colleagues and I ran in Mexico City (Bernard et al. 1989). At the time, we were just beginning to develop what became known as the network scale-up method for estimating the size of populations that you can't count. In this case, the population we wanted to estimate was the number of people who had died in the 1985 Mexico City earthquake. The codebook in table 19.1 contains four essential pieces of information:

1. The column number(s) in which each variable is coded, for each unit of analysis.
2. A short name for each variable.
3. A full, clear description of each variable.
4. A list of the possible values that each variable can take.

I will discuss these in turn.

1. *Column numbers.* Data are collected about units of analysis, like informants, songs, or court cases. Some studies, like this street-intercept survey, involve just a few variables; others may contain hundreds. Data used to be stored in lines of 80 spaces, or columns (this is because the first data entry machines mimicked typewriters, and there were 80 spaces on a normal page of type). In this case, we only needed 42 columns to hold the data for each respondent. In studies with more variables, one 80-column line wouldn't have been enough, so codebooks from data collected before about 1990 will have column numbers indicated. Contemporary data analysis programs have no such limitations

2. *Short name.* Make variable names as obvious as possible and keep them short. It is customary to begin each **data record** (that is, each line of data) with a **unique identifier** and to list that identifier as the first variable in a codebook. This usually means the informant number. We called the first variable here INTNUM, which was the interview number in the street-intercept project.

Notice that the variables in table 19.1 contain up to eight characters. At the time (1987), we had clunky computers and programs that could only handle eight-character names. Today, most data-analysis programs allow long variable names, but I still like short ones because they keep data columns narrow and you can see more of them on a single screen.

TABLE 19.1
Codebook for Mexico City Network Study, January 15, 1987

Column	Variable Name	Variable Description
1-4	INTNUM	Interview number, from 0001-2400
5-7	ZONE	Number of the zone of the city, from 1-120 on a map of Mexico City where interview was conducted
8	CLAZONE	Socioeconomic class of the zone of the city, as determined by the interviewers. 1 = Lower Class, 2 = Middle Class, 3 = Upper Class
9	SEX	The sex of the respondent. 1 = Male, 2 = Female
10	CLASRESP	Socioeconomic class of the respondent, as determined by the interviewers. 1 = Lower Class, 2 = Middle Class, 3 = Upper Class
11-12	AGE	Age of respondent, self-reported, in years
13-14	SCHOOL	Number of years respondent spent in school
15-16	DFLIVE	Number of years respondent has lived in Mexico City [D.F., or Distrito Federal]
17	OCCN	Occupation. 1 = Housewife, 2 = Regular employment, 3 = Retired, 4 = Unemployed, 5 = Other
18	DOC	Does the respondent know a physician who works in the public hospitals? 1 = Yes, 2 = No
19	YESDOC	Does the interviewer think the respondent really does know a physician in the public hospitals? 1 = Yes, 2 = No
20	QUAKE	Does the respondent know someone who died in the 1985 earthquake? 1 = Yes, 2 = No
21	YESQUAKE	Same as YESDOC, column 19, for QUAKE
22	HOWLONG	How many days did it take before the respondent learned that someone he or she knew had died in the quake? 1 = 0-15 days, 2 = 15-30 days, 3 = 31-45 days, 4 = 46-60 days, 5 = 61+ days
23	MAIL	Does the respondent know someone who works for the postal authority? 1 = Yes, 2 = No
24	YESMAIL	Same as YESDOC, column 19, for MAIL
25	ROBBED	Does the respondent know someone who was robbed in the street in 1986? 1 = Yes, 2 = No
26	YESROB	Same as YESDOC, column 19, for ROBBED
27	BUS100	Does the respondent know someone who is a bus driver on Route 100? [This is a public bus route in Mexico City.] 1 = Yes, 2 = No
28	YESBUS	Same as YESDOC, column 19, for BUS100
29	PESERO	Does the respondent know someone who drives a <i>pesero</i> in Mexico City? [These are private cars and Volkswagen minibuses that operate along established routes as privately owned, public conveyances.] 1 = Yes, 2 = No
30	YESPES	Same as YESDOC, column 19, for PESERO
31	PRIEST	Does the respondent know a Catholic priest in Mexico City? 1 = Yes, 2 = No
32	YESPRIES	Same as YESDOC, column 19, for PRIEST
33	VENDOR	Does the respondent know someone who works as a street vendor in the underground economy? 1 = Yes, 2 = No
34	YESVEND	Same as YESDOC, column 19, for VENDOR
35	TV	Does the respondent know someone who works for Televisa, the television company in Mexico City? 1 = Yes, 2 = No
36	YESTV	Same as YESDOC, column 19, for TV
37	WIND	Does the respondent know someone who makes his living cleaning car windshields at stoplights in Mexico City and asking for tips? 1 = Yes, 2 = No
38	YESWIND	Same as YESDOC, column 19, for WIND
39	RAPE	Does the respondent know someone who was raped in the street in 1986? 1 = Yes, 2 = No
40	YESRAPE	Same as YESDOC, column 19, for RAPE
41	INTID	Identity of the interviewer. 1 = Alicia, 2 = Juana, 3 = Roberto, 4 = Glauccio, 5 = Gladys, 6 = Alberto
42	KNOW	How many people does the respondent think he or she knows? 1 = 0-100, 2 = 101-500, 3 = 501-1,000, 4 = 1,001-1,500, 5 = >1,500

Here are some examples of common variable names you'll see in social research: AGE, INCOME, EDUC, HOUSETYP (house type), OWNCAR (does the informant own a car?), MIGRATE (does the informant have plans to migrate?), PQOL (perceived quality of life). It is customary to use all capital letters for variable names, but feel free to violate this and any other custom if it'll help you do the analysis. Ricardo Godoy, for example, a seasoned quantitative researcher in anthropology, regularly uses variable names like WEALTHTOTALAEQ to indicate "the value of household assets divided by the total number of adult equivalents in a household," where each child counts for a specified fraction of an adult equivalent in terms of calories per day consumed (<http://www.heller.brandeis.edu/sid/downloads/appendix.doc>).

Of course, each project is different and will contain variables that are specific to the research. Some examples that I've seen recently are VISNAI (has the informant ever visited Nairobi?), BIRTHCON (what is the informant's position on birth control?), and DISTH20 (how far is the household from potable water?).

3. *Variable description.* Be as clever as you like with variable names; just be sure to include a chatty description of each variable in the codebook so you'll know what all those clever names mean a year later! Leave nothing to the imagination of the user of a codebook (and that includes you). For example, "age of informant, to the nearest year, and reported by the informant," is much better than "age of informant."

Some variables require a lot of description. Consider this helpful codebook description of a variable: "Perceived Quality of Life. This was measured using an index consisting of the six items that follow. Each item is scored separately, but the items can be added to form an index. Since each item is scored from 1 to 5, the index of perceived quality of life can vary from 6 to 30 for any informant."

If you are using an established index or scale or data collection technique, then name it (i.e., "the Bogardus social distance scale") and provide a citation to the source (Bogardus 1933). If you have adapted a published technique to meet your particular needs, then mention that in your codebook, too. For example, "I have used the Fischer method of generating social networks (Fischer 1982) but have adapted it in translation for use with the Kipsigis."

Later, you or another researcher can compare the relevant items on your survey instrument with those in the published index. And always file a copy of any survey instrument with your codebook.

4. *Variable values.* Specify the values that each variable can take and what each value means. Marital status, for example, is often coded as 1 = married, 2 = divorced, 3 = separated, 4 = widowed, 5 = never married, 6 = unknown. Religion might be coded as 1 = Shintoist, 2 = Buddhist, 3 = Hindu, 4 = Tribal, 5 = Unknown.

Notice in table 19.1 that I violated my own rule, in *two* places, about not analyzing data until you have them. See those variables HOWLONG and KNOW? The codes for these variables were set during a brainstorming session in Mexico with the interview team after they'd done a few test interviews on the street. It turned out that people just couldn't answer those two questions (How many days was it before you learned that this person had died in the quake? How many people do you think you know, all told?) by giving a number, so the team came up with some chunks (31–45 days, 501–1,000 people, etc.) that they thought were reasonable.

And remember to mark **string variables** (also called **character variables**) in your codebook. String variables are alphabetic characters and have no numerical properties. If you have the name of the interviewer as a variable (in case you want to check for interviewer response effects), that name is a string of characters.

Cleaning Data

No matter how careful you are, you're going to make mistakes in coding and entering data into a computer. Even if you code behavioral data in the field directly into a handheld computer, believe me, you'll make mistakes. Everyone does. The trick is to catch the mistakes.

The best way to catch coding and data entry errors is with a data editor, like those that come with many statistical analysis packages (see appendix F). Suppose you have a variable called NUMCHILD, or "number of children." If you suspect that none of your informants have more than 10 children, you can tell the editor to "find the cases in the column for NUMCHILD where the entry is greater than 10." If the program finds such cases, you can decide whether the entry is an error or if it just means the informant has a lot of kids. Similarly, if an informant says that she has never had children, you want to make sure that the variable for number of grandchildren is also 0.

You can move columns of numbers around with a data editor (see appendix F). When you enter data, you often don't know which variables to put together. Later, as you become more familiar with your data, you'll want to move columns around so you can eyeball the data matrix and look for mistakes and incompatibilities. Once you have your data into the computer and cleaned, you're ready to start analyzing them.

Raw Data

The first thing to do, before you try any fancy statistical operations on your data, is lay them out and get a feel for them. How many cases are there of

people over 70? What is the average number of children in each household? How many people in your sample have extreme views on some key attitude questions?

Table 19.2 shows the raw data for five variables and 30 respondents. These data come from a telephone survey that Gery Ryan, Stephen Borgatti, and I did of 609 adults in the United States (Bernard et al. 1999). Part of the survey was about people's attitudes toward environmental activism. (The 30 respondents in table 19.2 are a random sample of the 609.)

The first variable, gender, is a nominal, or **qualitative variable**. The respondents were men and women over the age of 18, selected randomly from across the 48 continental states of the United States. Men were coded as 1 (GENDER = male); women were coded as 2 (GENDER = female). In statistics, qualitative description entails assigning numbers to classes of things. Those numbers, though—like 1 for male and 2 for female—are just substitute names for “male” and “female.” They are not quantities. The average of the 1s and 2s in the column for GENDER in table 19.2 is 1.4, but that's no more helpful than knowing the average telephone number in New York City.

The second two variables are items that Kempton et al. (1995) used in their study of environmental values in America. My colleagues and I wanted to see if we could replicate their results. These ordinal variables are responses, on a scale of 1 to 5, to two statements:

Americans are going to have to drastically reduce their level of consumption over the next few years.

- <1> strongly disagree
- <2> disagree
- <3> neutral
- <4> agree
- <5> strongly agree

Environmentalists wouldn't be so gung-ho if it were *their* jobs that were threatened.

- <1> strongly disagree
- <2> disagree
- <3> neutral
- <4> agree
- <5> strongly agree

I've labeled the responses to the two items REDUCE and GUNGHO in table 19.2. Notice that these two items are sort of opposites. The more you agree with REDUCE, the *stronger* your support for environmentalist issues is likely to be. But the more you agree with GUNGHO, the *weaker* your support for environmentalist issues is likely to be. If we want bigger numbers, like 4

TABLE 19.2
30 Records from Bernard et al.'s Study of Green Attitudes

<i>RESP</i>	<i>GENDER</i>	<i>REDUCE</i>	<i>GUNGHO</i>	<i>AGE</i>	<i>EDUC</i>
1	2	5	1	46	18
2	2	4	2	56	12
3	1	4	3	25	18
4	1	5	4	24	12
5	1	4	2	60	5
6	1	2	4	51	18
7	1	2	4	53	14
8	2	5	4	25	13
9	1	2	4	21	15
10	2	4	4	67	13
11	2	2	1	34	16
12	1	5	3	47	18
13	1	4	1	35	12
14	2	4	3	67	12
15	1	4	4	20	12
16	2	4	2	24	15
17	1	5	2	38	16
18	2	5	2	53	14
19	1	4	2	38	12
20	2	4	3	31	14
21	1	5	4	54	15
22	2	5	4	52	14
23	1	4	2	37	14
24	1	2	1	53	14
25	1	5	5	49	18
26	1	3	3	46	16
27	2	5	4	78	14
28	2	5	2	41	12
29	1	4	2	57	12
30	1	4	4	69	10

and 5, always to stand for support of environmentalism and smaller numbers, like 1 and 2, always to stand for lack of support, then we have to transform the data for GUNGHO so that they run in the same direction as the data for REDUCE. We can easily do that in any statistics package.

This is a very important part of univariate analysis—getting the data into the computer in just the right form.

Variables 4 and 5, AGE (the respondent's age) and EDUC (the respondent's

level of education), are interval. (They are really ratio variables, but recall from chapter 2 that ratio variables are conventionally referred to as “interval.”) For AGE, we simply asked respondents “How old are you?” Here is the question from the survey that produced the data for EDUC:

What is the highest grade of school or year in college you yourself completed?

None	0
Elementary	01
Elementary	02
Elementary	03
Elementary	04
Elementary	05
Elementary	06
Elementary	07
Elementary	08
High School	09
High School	10
High School	11
High School	12
College—one year	13
College—two years	14
College—three years	15
College—four years	16
Some Graduate School	17
Graduate/Prof. Degree	18

Frequency Distributions

Tables 19.3a–e show the raw data from table 19.2 transformed into a set of **frequency distributions**. I used SYSTAT® to produce tables 19.3a–e, but any program will do.

TABLE 19.3a
Frequency Table of the Variable GENDER

<i>Count</i>	<i>Cum. Count</i>	<i>Percent</i>	<i>Cum. Percent</i>	<i>Variable GENDER</i>
18	18	60	60	Male
12	30	40	100	Female

TABLE 19.3b
Frequency Table of the Variable AGE

<i>Count</i>	<i>Cum. Count</i>	<i>Percent</i>	<i>Cum. Percent</i>	<i>Variable AGE</i>
1	1	3.3	3.3	20
1	2	3.3	6.7	21
2	4	6.7	13.3	24
2	6	6.7	20.0	25
1	7	3.3	23.3	31
1	8	3.3	26.7	34
1	9	3.3	30.0	35
1	10	3.3	33.3	37
2	12	6.7	40.0	38
1	13	3.3	43.3	41
2	15	6.7	50.0	46
1	16	3.3	53.3	47
1	17	3.3	56.7	49
1	18	3.3	60.0	51
1	19	3.3	63.3	52
3	22	10.0	73.3	53
1	23	3.3	76.7	54
1	24	3.3	80.0	56
1	25	3.3	83.3	57
1	26	3.3	86.7	60
2	28	6.7	93.3	67
1	29	3.3	96.7	69
1	30	3.3	100.0	78

TABLE 19.3c
Frequency Table of the Variable EDUC

<i>Count</i>	<i>Cum. Count</i>	<i>Percent</i>	<i>Cum. Percent</i>	<i>Variable EDUC</i>
1	1	3.3	3.3	5
1	2	3.3	6.7	10
8	10	26.7	33.3	12
2	12	6.7	40.0	13
7	19	23.3	63.3	14
3	22	10.0	73.3	15
3	25	10.0	83.3	16
5	30	16.7	100.0	18

TABLE 19.3d
Frequency Table of the Variable REDUCE

<i>Count</i>	<i>Cum. Count</i>	<i>Percent</i>	<i>Cum. Percent</i>	<i>Variable REDUCE</i>
5	5	16.7	16.7	2
1	6	3.3	20.0	3
13	19	43.3	63.3	4
11	30	36.7	100.0	5

TABLE 19.3e
Frequency Table of the Variable GUNGHO

<i>Count</i>	<i>Cum. Count</i>	<i>Percent</i>	<i>Cum. Percent</i>	<i>Variable GUNGHO</i>
4	4	13.3	13.3	1
9	13	30.0	43.3	2
5	18	16.7	60.0	3
11	29	36.7	96.7	4
1	30	3.3	100.0	5

Using a Distribution Table

One thing to look for in a frequency distribution is variability. If a variable has no variability, then it is simply not of any further interest. If everyone in this sample of respondents were the same gender, for instance, we wouldn't use GENDER in any further analysis. Looking carefully at the frequency distribution is your first line of defense against wasting a lot of time on variables that don't vary.

We see from table 19.3a that 60% of the sample are men. Table 19.3b shows that age is pretty evenly distributed. Table 19.3c shows that two-thirds of the sample (20 out of 30) had more than a high school education and that 5 of the 30 had graduate degrees. Most people (24 out of 30) agreed or strongly agreed with the statement that Americans are going to have to reduce consumption drastically in the coming years (table 19.3d).

People were pretty evenly split, though, on whether environmentalists would be so gung-ho if their jobs were threatened (table 19.3e): 13 people either disagreed or strongly disagreed with that sentiment, and 12 people either agreed or strongly agreed (5 were neutral).

Frequency distributions give you hints about how to collapse variables. With 609 respondents in the full survey, we had plenty of responses to all

possible answers for the question about reducing consumption. But with a sample of just 30 responses in table 19.2 above, we didn't get anyone who said they strongly disagreed with the statement that Americans are going to have to drastically reduce their consumption in the coming years.

You can see this in the far right-hand column of table 19.3d. If all we had were these 30 cases, we'd want to create a three-category variable—disagree, neutral, and agree—by collapsing the data for REDUCE into (1) the 5 people who answered 2 (disagree), (2) the 1 person who answered 3 (neutral), and (3) the 24 people who answered 4 or 5 (agree or strongly agree).

It's not always obvious how to group data. In fact, it's often better not to. Look at table 19.3c. Only two people in our sample of 30 had less than 12 years of education. We could conveniently group those two people into a category called "less than high school." There is a bulge of eight people who had 12 years of education (they completed high school), but then we see just two people who reported a year of college and seven people who reported 2 years of college. That bulge of seven respondents might be people who went to a community college. We might group those two sets of people into a category called "up to two years of college."

Those three people in table 19.3c who reported 4 years of college form an obvious class ("finished college"), and so do the five people who reported having a graduate or professional degree that required more than 4 years of college. But what do we do with those three people who reported 3 years of college? We could lump them together with the three respondents who finished college, but we could also lump them with the nine people who reported 1 or 2 years of college.

The problem is, we don't have any ironclad decision rule that tells us how to lump data into categories. We don't want to maintain a separate category of just three respondents (the people who reported 3 years of college), but we don't know if they "belong" (in some socially important sense) with those who had some college or with those who completed college.

I recommend *not* grouping interval-level data unless you really have to. No matter which decision you make about those three people who reported 3 years of college in table 19.3c, you're turning an interval-level variable (years of education) into an ordinal-level variable (less than high school, high school, etc.). As you saw earlier, with the variables HOWLONG and KNOW in the Mexico City study, there are times when this is a good idea, but trading interval for ordinal measurement means throwing away data. You need a really good reason to do that.

Measures of Central Tendency

Once we have the data laid out and have a feel for what's in there, we can start describing the variables. The first thing to do is get some overall measure of the “typical” value for each variable. This is called a measure of **central tendency**.

The three most widely used measures of central tendency are the **mode**, the **median**, and the **mean**. All these get packaged together in everyday speech as some kind of “average,” but we have to be more precise in data analysis. Each measure of central tendency carries important information about the values of a variable.

Here are the definitions for each of these measures of central tendency:

1. *The mode is the attribute of a variable that occurs most frequently.* The mode can be found for nominal-, ordinal-, and interval-level variables, but it is the only measure of central tendency available for nominal variables.
2. *The median is the midpoint in a distribution above and below which there are an equal number of scores in a distribution.* The median can be found for ordinal- and interval-level variables.
3. *The mean, or the average, is the sum of the individual scores in a distribution, divided by the number of scores.* The mean can be found for ordinal- and interval-level variables.

Central Tendency: The Mode

The mode is the attribute of a variable that occurs most frequently. Technically, the mode is not calculated; it is observed. You find it by simply looking at the data and seeing which attribute of a variable occurs the most. In table 19.3b, we see that the modal value for education is 12 years (there are 8 out of 30 cases). This tells us that finishing high school is the most common level of education in our sample of 30 respondents.

All variables (nominal, ordinal, and interval) have modal values, but nominal variables can *only* have modal values. In table 19.3a, for example, we see that there are 18 men (GENDER = 1) and 12 women (GENDER = 2). The mode, then, is male for this sample of 30. (The mode, by the way, was female for the full survey of 609 respondents. When you work with small samples, fluctuations of this magnitude are normal.)

Many distributions have more than one mode, and bimodal distributions are quite common. In a rural community that has experienced a lot of out-migration, for example, the age structure is likely to be bimodal: There are young

people hanging around who aren't old enough to leave and old people who can't find work in the city because of their age.

Using the Mode

The mode is often said to be the weakest measure of central tendency, but it's very useful when you want to make a statement about a prominent qualitative attribute of a group. "More people profess to be Buddhist in this prefecture of Japan than profess any other religion" is such a statement.

The mode is also a good common-sense alternative to the sometimes unrealistic quality of the mean. Saying that "the modal family size is 4 people" makes a lot more sense than saying that "the average family size is 3.81 people"—even if both statements are true.

You'll often see the mode reported as a percentage: "In this survey, 60% of the respondents were men."

The mode can also be reported in terms of **ratios**. Of the 30 respondents in table 19.2, 12 were women and 18 were men, so the modal value for the variable GENDER in table 19.2 is male. The ratio of men to women among these respondents is $18/12 = 1.5$, while the ratio of women to men is $12/18 = .67$. Reporting that "there were 1.5 men for every woman in this survey" is the same as saying that "60% of the respondents were men." Reporting that "there were .67 women for every man in this survey" is the same as saying that "40% of the respondents were women."

Central Tendency: The Median

The median is the point in a distribution above and below which there are an equal number of scores in a distribution. If you've ever taken a standardized test like the ACT, SAT, or GRE, you might have been told that you scored in the 14th percentile, or the 31st percentile, or whatever. If you scored in the 86th percentile, then 14% of the scores were higher than yours ($.86 + .14 = 1.0$).

Ten percent of scores in a list are below the 10th percentile, and 90% are above it. The 25th percentile is called the **first quartile** and the 75th percentile is the **third quartile**. The difference between the values for the 25th and 75th percentiles is known as the **interquartile range** and is a measure of dispersion for ordinal and interval-level variables. (More on measures of dispersion later.)

The median is the 50th percentile. It can be used with ranked or ordinal data and with interval- or ratio-level data. For an *odd number* of unique obser-

vations on a variable, the median score is $(n + 1)/2$, where n is the number of cases in a distribution and the scores are arranged in order.

Suppose we ask nine people to tell us how many brothers and sisters they have, and we get the following answers:

0 0 1 1 1 1 2 2 3

The median observation is 1 because it is the middle score—there are four scores on either side of it, $(n + 1)/2 = 5$, and we see that the median is the fifth case in the series, once the data are arranged in order.

Often as not, of course, as with the data on those 30 respondents from the green survey shown in table 19.2, you'll have an *even number* of cases. Then the median is the average of $n/2$ and $(n/2 + 1)$, or the midpoint between the *two* middle observations, once the data are arranged in order. I asked 16 undergraduate students "How long do you think you'll live?" Here are the responses:

70 73 75 75 79 80 80 83 85 86 86 87 87 90 95 96

$n/2 = 8$ and $n/2 + 1 = 9$, so the median is 84, midway between the two middle observations, 83 and 85. (By the way, if the two middle observations had been, say, 83, then the midpoint between them would be 83.)

The Median of Grouped Data

A lot of data are reported in intervals, or groups. For example, some people are uncomfortable with a straightforward question like "How much money do you make?" so researchers often ask something like:

Now we'd like to get an idea of about how much you earn each year. Do you earn:

- (1) less than \$10,000 per year?
- (2) \$10,000 or more but less than \$20,000 per year?
- (3) \$20,000 or more but less than \$30,000 per year?
- (4) \$30,000 or more but less than \$40,000 per year?
- (5) \$40,000 or more but less than \$50,000 per year?

and so on. This produces **grouped data**. In table 19.3b, the two middle scores for AGE are 46 and 47, so the median is 46.5. Table 19.4 shows the data on AGE from table 19.3b, grouped into 10-year intervals. To find the median in grouped data, use the formula for finding any percentile score in a distribution of grouped data:

TABLE 19.4
Frequency Table of the Grouped Variable AGE

<i>Count</i>	<i>Cum. Count</i>	<i>Variable AGE</i>
6	6	20–29
6	12	30–39
5	17	40–49
8	25	50–59
5	30	60+

$$PS = L + i \left\{ \frac{n - C}{f} \right\} \quad \text{Formula 19.1}$$

where:

- PS is the percentile score you want to calculate;
- L is the true lower limit of the interval in which the percentile score lies;
- n is the case number that represents the percentile score;
- C is the cumulative frequency of the cases up to the interval *before* the one in which the percentile score lies;
- i is the interval size; and
- f is the count, or *frequency*, of the interval in which the median lies.

In applying formula 19.1 to the data in table 19.4, the first thing to do is calculate n . There are 30 cases and we are looking for the score at the 50th percentile (the median), so n is $(30)(.50) = 15$. We are looking, in other words, for a number *above which* there are 15 cases and *below which* there are 15 cases. Looking at the data in table 19.4, we see that there are 12 cases up to 39 years of age and 17 cases up to 49 years of age. So, C is 12, and the median case lies somewhere in the 40–49 range.

The true lower limit, L, of this interval is 39.5 (midway between 39 and 40, the boundary of the two groups), and the interval, i , is 10 years. Putting all this into the formula, we get:

$$PS = 39.5 + 10 \left\{ \frac{15 - 12}{5} \right\} = 39.5 + 10(.6) = 45.5$$

So, the median age for the grouped data on 30 people is 45.5 years. Notice that: (1) The median for the *ungrouped* data on AGE in table 19.3b is 46.5, so in grouping the data we lose some accuracy in our calculation; and (2) None

of the respondents in table 19.3b actually reported a median age of 45.5 or 46.5.

I've given you this grand tour of the median as a specific percentile score because I want you to understand the conceptual basis for this statistic. I'm going to do the same thing for all the statistical procedures I introduce here and in the next two chapters. You only need to work through these detailed examples once. When you understand the concepts behind the median, the standard deviation, z -scores, chi-square, t -tests, and regression, you need never again calculate these statistics by hand. Once you understand a statistic, you can do all the calculations by computer.

In fact, even for small data sets, it's best to use a computer to handle statistical chores. It's not just easier to do—it's more accurate. You're less likely to make mistakes in recording data on a computer and when you do make mistakes (I make them all the time), it's easier to find and fix them. Just think of how easy it is to find spelling errors in a document when you use a word processor. It's the same with checking and correcting mistakes in statistical analysis.

Central Tendency: The Mean

The arithmetic mean, or the average, is the *sum of the individual scores in a distribution, divided by the number of scores*. The formula for calculating the mean is:

$$\bar{x} = \sum x / n \quad \text{Formula 19.2}$$

where \bar{x} (read: x -bar) is the mean, $\sum x$ means "sum all the values of x " and n is the number of values of x . To calculate the mean, or average age of the 30 respondents whose data are shown in table 19.2, we add up the 30 ages and divide by 30. The mean age of these 30 respondents is 45.033 years. (We use \bar{x} when we refer to the mean of a sample of data; we use the Greek letter μ when we refer to the mean of an entire population.)

The formula for calculating the *mean of a frequency distribution* is:

$$\bar{x} = \sum fx / n \quad \text{Formula 19.3}$$

where $\sum fx / n$ is the sum of the attributes of the variable, times their frequencies.

Table 19.5 shows the calculation of the mean age for the frequency distribution shown in table 19.3b.

TABLE 19.5
Calculating the Mean for the Data in Table 19.3b

Count (<i>f</i>)	AGE (<i>x</i>)	<i>fx</i>
1	20	20
1	21	21
2	24	48
2	25	50
1	31	31
1	34	34
1	35	35
1	37	37
2	38	76
1	41	41
2	46	92
1	47	47
1	49	49
1	51	51
1	52	52
3	53	159
1	54	54
1	56	56
1	57	57
1	60	60
2	67	134
1	69	69
1	78	78
		$\sum fx = 1,351$
		$\sum fx/n = 1,351/30 = 45.033$

Calculating the Mean of Grouped Data

Table 19.6 shows the calculation of the mean for the grouped data on AGE in table 19.4. When variable attributes are presented in ranges, as in the case here, we take the midpoint of the range.

Note the problem in taking the mean of the grouped data in table 19.6. If you go back to table 19.3b, you'll see that all six of the people who are between 20 and 29 are really between 20 and 25. Counting them all as being 25 (the midpoint between 20 and 29) distorts the mean.

Also, there are five people over 60 in this data set: one who is 60, two who are 67, and one each who are 69 and 78. *Their* average age is 68.2, but they are all counted as being just 60+ in table 19.4. In calculating the mean for

TABLE 19.6
Calculating the Mean for the Frequency Table of the Grouped Variable AGE

<i>AGE range</i>	<i>midpoint (x)</i>	<i>f</i>	<i>fx</i>
20–29	25	6	150
30–39	35	6	210
40–49	45	5	225
50–59	55	8	440
60+	65	5	325
		$n = 30$	$\sum fx = 1,350$
			$\bar{x} = 1,350/30$
			$= 45.00$

these grouped data, I've assigned the midpoint to be 65, *as if the range were 60–69*, even though the actual range is 60–78, and the real midpoint of the 60+ category is

$$(60 + 67 + 67 + 78)/5 = 68.2$$

If you have grouped data, however, it will usually be because the data were collected in grouped form to begin with. In that case, there is no way to know what the real range or the real midpoint is for the 60+ category, so we have to assign a midpoint that conforms to the midpoints of the other ranges. Applying the midpoint for all the other classes, the midpoint for the 60+ category would be 65, which is what I've done in table 19.6.

You can see the result of all this distortion: The grouped data have a mean of 45.00, while the ungrouped data have a calculated mean of 45.033. In this case, the difference is teeny, but it won't always be that way. If you collect data in groups about interval variables like age, you can never go back and see how much you've distorted things. It's always better to collect interval data at the interval level, if you can, rather than in grouped form. You can group the data later, during the analysis, but you can't "ungroup" them if you collect data in grouped form to begin with.

A Mathematical Feature of the Mean

The arithmetic mean has an important feature: *The sum of the deviations from the mean of all the scores in a distribution is zero.* While the median and the mean are both midpoints in sets of scores, the mean is the point in a distribution at which the two halves balance each other out. You can see this in table 19.7, which shows data from 10 countries on the percentage of women

TABLE 19.7

Data on Female Illiteracy in 10 Countries (from Table 19.8) Showing How the Sum of the Deviations from the Mean of a Distribution Is Zero

<i>COUNTRY</i>	<i>FEMILLIT</i> x	<i>Score - Mean</i> $x - \bar{x}$
Armenia	2.4	$2.4 - 19.49 = -17.09$
Chad	66.0	$66.0 - 19.49 = 46.51$
El Salvador	23.8	$23.8 - 19.49 = 4.31$
Ghana	37.1	$37.1 - 19.49 = 17.61$
Iran	30.1	$30.1 - 19.49 = 10.61$
Latvia	0.2	$0.2 - 19.49 = -19.29$
Namibia	18.8	$18.8 - 19.49 = -0.69$
Panama	8.7	$8.7 - 19.49 = -10.79$
Slovenia	0.4	$0.4 - 19.49 = -19.09$
Suriname	7.4	$7.4 - 19.49 = -12.09$
	$\sum x = 194.9$	$\sum (x - \bar{x}) = 0.00$
	$\bar{x} = 194.9/10 = 19.49$	
Scores below the mean		Scores above the mean
	-17.09	46.51
	-19.29	17.61
	-0.69	4.31
	-10.79	10.61
	-19.09	
	-12.09	
Total deviation below the mean	-79.04	Total deviation above the mean 79.04

over the age of 15 who are illiterate. This feature of the mean figures prominently in the calculation of variance, which is coming up in the section on measures of dispersion.

The data in table 19.7 are a random sample of the data in table 19.8. That table shows some **social indicators** for 50 randomly selected countries around the world. We'll be using the data in table 19.8 as we go along in the next few chapters.

The Outlier Problem

The mean is one of the all-time great statistics, but it has one very important drawback: In small samples, it's heavily influenced by special cases, called

TABLE 19.8
Some Social Indicators for 50 Randomly Selected Countries

<i>COUNTRY</i>	<i>LEXMALE</i>	<i>LEXFEM</i>	<i>INFMORT</i>	<i>MFRATIO</i>	<i>TFR</i>	<i>FEMILLIT</i>	<i>PCGDP</i>
Albania	69.9	75.9	30	105	2.50	23.0	972
Argentina	69.6	76.8	22	96	2.62	3.2	8,257
Armenia	67.2	73.6	26	95	1.70	2.4	533
Austria	73.7	80.2	6	97	1.41		25,911
Bahamas	70.5	77.1	15	97	2.60	3.6	11,395
Brunei	73.4	78.1	10	110	2.80	11.9	13,719
Cameroon	53.4	56.0	74	99	5.30	30.0	702
Cape Verde	65.5	71.3	56	88	3.56	33.8	1,085
Chad	45.7	48.7	112	98	6.07	66.0	150
Chile	72.3	78.3	13	98	2.44	4.5	4,921
China	67.9	72.0	41	106	1.80	23.7	777
Congo D.R.	49.2	52.3	90	98	6.43	49.8	98
Ecuador	67.3	72.5	46	101	3.10	10.5	1,620
El Salvador	66.5	72.5	32	96	3.17	23.8	1,941
Equatorial Guinea	48.4	51.6	108	97	5.58	25.6	377
Gambia	45.4	48.6	122	98	5.20	70.6	355
Ghana	58.3	61.8	66	99	5.15	37.1	346
Hungary	66.8	74.9	10	92	1.37	0.8	4,644
Iran	68.5	70.0	35	103	2.80	30.1	2,850
Ireland	73.6	79.2	7	99	1.90		23,098
Japan	76.8	82.9	4	96	1.43		29,956
Kenya	51.1	53.0	66	100	4.45	24.0	373
Laos	52.0	54.5	93	102	5.75	66.8	250
Latvia	62.4	74.4	18	83	1.25	0.2	2,638
Lesotho	54.7	57.3	93	97	4.75	6.4	425
Liberia	46.1	48.5	116	101	6.30	62.3	285
Maldives	65.7	63.3	50	106	5.40	3.6	1,350
Mongolia	64.4	67.3	51	101	2.60	0.7	384
Namibia	51.8	53.0	65	100	4.90	18.8	1,834
Nepal	57.6	57.1	83	103	4.45	76.1	197
Netherlands	75.0	80.7	6	98	1.50		24,956
Nicaragua	65.8	70.6	43	99	4.42	29.8	442
Pakistan	62.9	65.1	74	107	5.02	68.9	458
Panama	71.8	76.4	21	102	2.63	8.7	3,287
Poland	68.2	76.9	15	95	1.53	0.3	4,096
Romania	66.2	73.9	23	96	1.17	2.80	1,698
Samoa	69.3	73.6	23	109	4.15	21.0	1,255
Senegal	50.5	54.2	63	99	5.57	72.3	518
Slovenia	70.6	78.2	7	95	1.26	0.4	9,798
South Africa	51.5	58.1	59	96	3.25	15.4	3404
Spain	74.5	81.5	7	96	1.15	3.2	14,690
Sri Lanka	70.9	75.4	18	98	2.10	11.0	851
Sudan	53.6	56.4	71	101	4.61	53.7	305
Suriname	67.5	72.7	29	99	2.21	7.4	2,454
Switzerland	75.4	81.8	6	98	1.47		35,910

TABLE 19.8 (Continued)

COUNTRY	LEXMALE	LEXFEM	INFMORT	MFRATIO	TFR	FEMILLIT	PCGDP
Tanzania	46.8	49.1	81	98	5.47	32.9	213
Togo	47.6	50.1	84	98	6.05	59.2	344
United Arab Em.	73.9	76.5	16	172	3.42	21.1	19,506
Uruguay	70.4	78.0	18	94	2.40	1.8	6,333
Zambia	39.5	40.6	82	98	5.55	28.6	413

LEXMALE and LEXFEM are the life expectancies, at birth, for children born during 1995–2000. INFMORT is the number of children who die, per 1,000 born, during the first year of life. SOURCE: Population Division and Statistics Division of the UN Secretariat. *World Population Prospects: The 1998 Revision*, Vol. I: Comprehensive Tables; *Demographic Yearbook 1997*; and *Population and Vital Statistics Report*, Statistical papers, Series A Vol. LIII, No. 3. <http://www.un.org/popin/wdtrends.htm>

MFRATIO is the number of males per 100 females in 2000. SOURCE: calculated by the Statistics Division of the UN Secretariat from *Sex and Age Annual 1950–2050* (The 1998 Revision) (data on diskettes), supplemented by *Demographic Yearbook 1997*.

TFR, or total fertility rate, is an estimate of the number of children a girl born during 1995–2000 will bear if current fertility patterns hold and she lives through her entire child-bearing years. TFR is estimated by the Population Division of the UN Secretariat. <http://www.who.int/whr/en>. TFR90 is the total fertility rate for 1990. SOURCE: World Health Report 2000

FEMILLIT is the percentages of women, >15 years of age who are illiterate. SOURCE: UNESCO Institute for Statistics, assessment of February 2000, supplemented by *UNESCO Statistical Yearbook 1999*. <http://www.uis.unesco.org/en/stats/statistics/yearbook/YBIndexNew.htm>

PCGDP is the per capita gross domestic product in 2000, calculated in U.S. dollars by the Statistics Division of the UN Secretariat. <http://www.un.org/esa>

outliers, and even in large samples, it's heavily influenced by big gaps in the distribution of cases. The median, however, is not as strongly affected by extreme scores. You can see this in table 19.9.

TABLE 19.9
Comparing the Mean and the Median

	LEXFEM	LEXMALE
All 50 countries in Table 19.8		
Median	72.250	66.350
Mean	67.050	62.552
Without Chad, Liberia, and Gambia		
Median	72.500	66.800
Mean	68.228	63.626

Using the data in table 19.8 as a proxy for the 200-or-so countries of the world, the mean life expectancy for women in the world is 67.05 years and for men it's 62.55 years. If we remove the three cases in table 19.8 that have infant

mortality rates above 110 per 1,000—Chad, Liberia, and Gambia—the mean life expectancy for women rises 1.8% to 68.23 years, and for men it rises 1.7% to 63.63 years. Chad, Liberia, and Gambia comprise just 0.002 (one-fifth of 1%) of the world’s population, but the data from those countries pull down the mean life expectancy for women and men by a factor of about 9 (that is, $0.002\% \times 9 = 1.8\%$). *Median* life expectancy, however, rises just 0.0025 (four-tenths of 1%) for women when we remove Chad, Liberia, and Gambia, and it rises just 0.007 (seven-tenths of 1%) for men.

When interval data (like life expectancy) are normally distributed, the mean is the best indicator of central tendency. When interval data are highly skewed, the median is often a better indicator of central tendency. You absolutely must get a feel for the **shape of distributions** to understand what’s going on.

Shape: Visualizing Distributions

A really good first cut at understanding whether data are normal or skewed is to lay them out graphically. This is easy to do with any of the full-featured statistics programs out there these days. I’ll show you six ways to lay out your data: bar graphs and pie charts for nominal and ordinal variables; stem-and-leaf plots, box-and-whisker plots, histograms, and frequency polygons for interval variables.

Bar Graphs and Pie Charts

Bar graphs and **pie charts** are two popular ways to graph the distribution of nominal and ordinal variables. Figure 19.1 shows bar graphs for two of the variables in table 19.2: GENDER and GUNGHO. Figure 19.2 shows the pie charts for the same variables. Notice that in the bar charts, the bars don’t touch one another. This indicates that the data are nominal or ordinal and not continuous.

The categories of the variables are shown along the horizontal axis of the bar graph. The horizontal, or *x*-axis, is also called the **abscissa**. The number of each category is shown on the left vertical axis. The vertical, or *y*-axis, is also called the **ordinate**. You can, of course, show the percent of each category on the *y*-axis instead.

In figure 19.1a, men are labeled 1 and women are labeled 2. Notice that it makes no difference whether we put the bar for men or the bar for the women on the left or the right when we graph GENDER. There is no order implied in the attributes of a nominal variable. When we graph ordinal variables, however, like GUNGHO, the order of the bars becomes important.

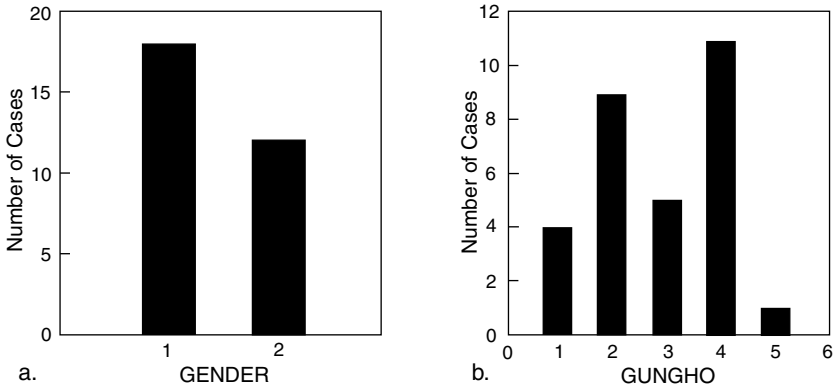


Figure 19.1. Bar charts for the variables GENDER and GUNGHO in Table 19.2.

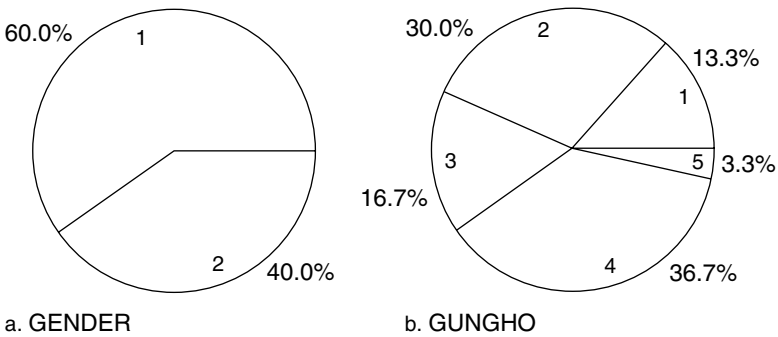


Figure 19.2. Pie charts for the variables GENDER and GUNGHO in Table 19.2.

Stem-and-Leaf Plots

I like to start visualizing interval data by running stem-and-leaf plots and box plots. Figure 19.3 shows a stem-and-leaf plot for the variable MFRATIO in table 19.8.

Inspection of table 19.8 confirms that the lowest value for this variable is 83 men per 100 women (in Latvia). The “stem” in the stem-and-leaf plot is 83, and the 0 next to it is the “leaf.” There are three cases of 95 men per 100 women (Poland, Slovenia, and Armenia), so there are three 0s after the 95 stem. The M in figure 19.3 stands for the median, or the 50th percentile (it’s 98), and the Hs indicate the upper hinge and the lower hinge, or the 25th and the 75th percentiles. The two cases below 92 and the three cases above 107 are

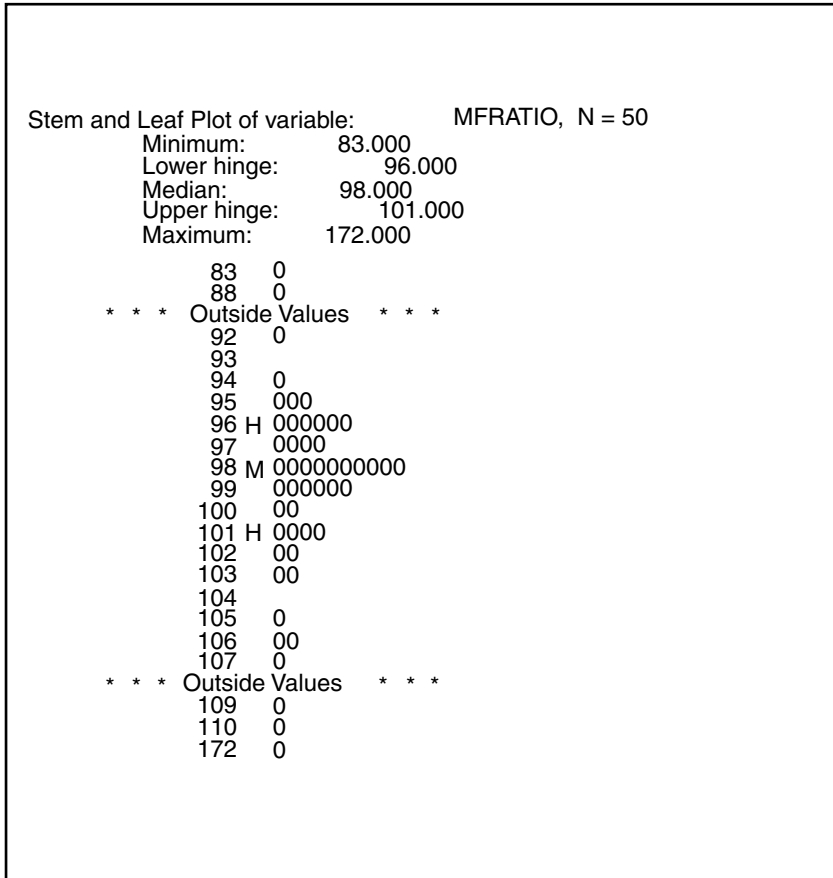


Figure 19.3. Stem-and-leaf plot for the variable MFRATIO in Table 19.8.

technically considered outliers. This variable looks like it's pretty normally distributed, with a hump of cases in the middle of the distribution and tails going off in both directions.

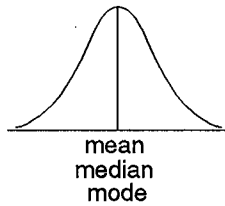
Distributions Again

We can check to see if the variable is, in fact, normally distributed or skewed by evaluating the distribution. That's where box plots come in.

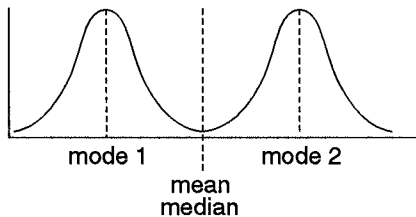
Just as a refresher, back in chapter 7, on sampling theory, we looked at

some basic *shapes* of distributions: skewed to the right, skewed to the left, bimodal, and normal. I have shown them again in figure 19.4, but this time notice the mode, the median, and the mean:

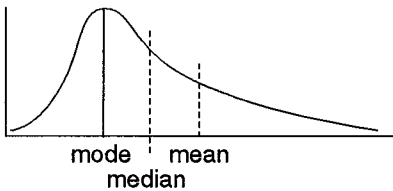
1. the mode, the median, and the mean are the same in the normal distribution;
2. the mean is pulled to the left of the median in negatively skewed distributions;
3. the mean is pulled to the right of the median in positively skewed distributions;
- and
4. some distributions are bimodal, or even multimodal. In these cases, you need to calculate two or more means and medians because the global mean and median give you a distorted understanding of the distribution.



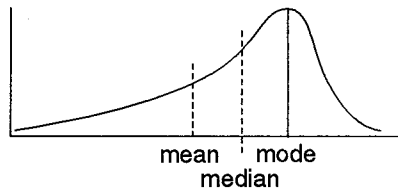
a. Normal distribution



b. Bimodal distribution



c. Skewed to the right



d. Skewed to the left

Figure 19.4. Some basic types of distributions.

Box-and-Whisker Plots

Stem-and-leaf plots give you exact values for the median, the hinges of the interquartile range, and the cases in the tails of a distribution. Box-and-whisker plots (or, simply, **box plots**) throw out the details and show you just the shape of a distribution. Together, the quantitative data from a stem-and-leaf plot and the purely qualitative image of a box plot give you a lot of useful information.

Figure 19.5 shows the box plots for MFRATIO (with and without the outliers), PCGDP and TFR. The boxes themselves show you the interquartile

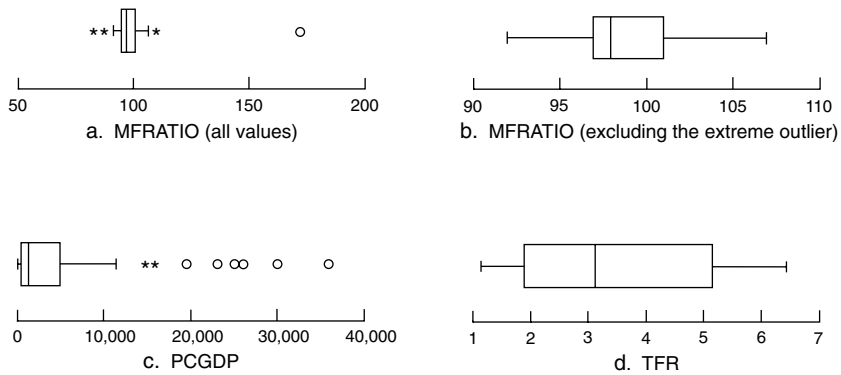


Figure 19.5. Box-and-whisker plots for four social indicators of 50 countries (from Table 19.8).

range—the middle 50% of the cases. The vertical line that marks off the box at the left is the 25th percentile (the lower hinge) of the plot; the vertical line that marks off the box at the right is the 75th percentile (the upper hinge). The vertical line inside the box is the median, or the 50th percentile. Fifty percent of the cases in the distribution fall to the right of the median line and 50% fall to the left.

The whiskers in box plots extend *one-and-a-half times the interquartile range* from the lower and upper hinges of the box. When data are normally distributed, this will be about 2.7 standard deviations from the mean. Cases outside that range are outliers and are marked by an asterisk. Cases that are more than *three times the interquartile range* from the hinges are extreme outliers and are marked by little circles.

The box plot for MFRATIO (figure 19.5a) shows that, in 50% of the countries of the world, the interquartile range is a very, very narrow band, with very, very short whiskers, a few outliers, and one extreme outlier. In fact, the

interquartile range for this variable is between 96 and 100 men per 100 women. The mean is 100.2 and the median is 98.0—slightly skewed to the left but quite close to a normal distribution.

Figure 19.5b shows MFRATIO with the extreme outlier excluded. Now the mean and the median for this variable are almost identical (98.0 and 98.8) and the distribution looks more-or-less normal. This is not a case for throwing out outliers. It *is* a case for examining every variable in your data set to get a feel for the distributions as the first step in analysis.

Figure 19.5c shows that 50% of the countries in the world had a very low per capita gross national product. There is almost no whisker to the left of the box, indicating that only a very few countries were below the interquartile range. There is a long whisker to the right of the box, as well as two outliers (asterisks) and six extreme outliers (circles). The circle farthest to the right is Switzerland, with over \$35,000 in per capita GNP. To put this in perspective, Luxembourg had the world's highest PCGDP in 2000, at \$43,475. The United States was seventh, at \$31,746. Mozambique was at the bottom, at \$92. Mozambique came up to \$1,200 in 2004, but by then, the United States had gone up to nearly \$38,000 (<http://www.odci.gov/cia/publications/factbook/>). The shape of the box-and-whisker plot in figure 19.5c reflects the fact that the world's population is divided between the haves and the have-nots.

Finally, figure 19.5d shows that total fertility in 2000 ranged between about 2 and about 5, and that there were no outliers (that is, no cases beyond the whiskers). This is another variable that looks like it might be normally distributed, but notice that the median (the midline in the box) is off center to the left. The mean for TFR is 3.14, lower than the median, which is 3.47. We need more information. Read on.

Histograms and Frequency Polygons

Two other graphic methods are useful in univariate analysis: **histograms** and **frequency polygons** (also called **frequency curves**). Frequency polygons are line drawings made by connecting the tops of the bars of a histogram and displaying the result without the bars. What you get is pure shape—less information than you get with box plots or stem-and-leaf plots, but easily interpreted. Figure 19.6 shows the histograms and frequency polygons for MFRATIO, PCGDP, and TFR.

The histogram and frequency polygon for MFRATIO nail down what we know from the box-plot of that variable: Most countries are in a narrow range for this variable, but a few are real outliers. The histogram/polygon for PCGDP removes all doubt about the shape of this variable: It is very highly skewed to the right, with most countries at the bottom and the long tail of

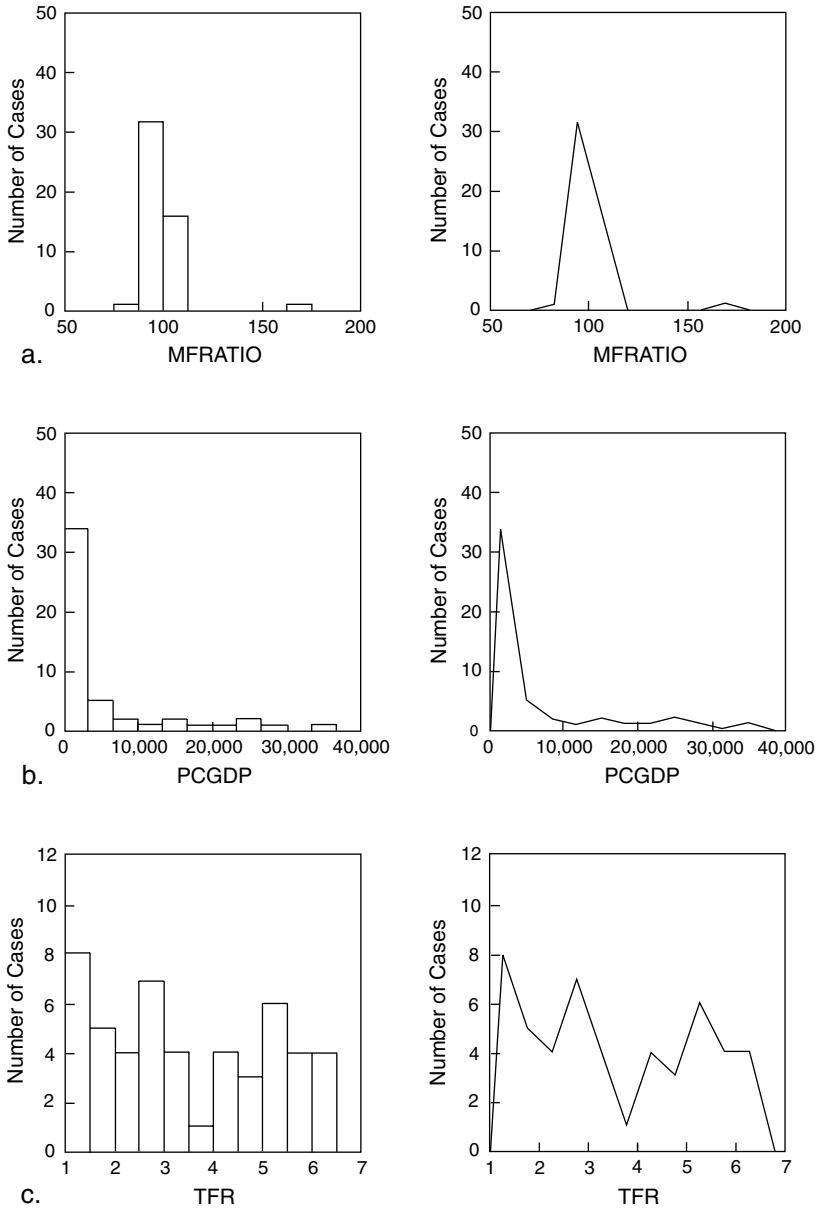


Figure 19.6. Histograms (*left*) and frequency polygons (*right*) for MFRATIO, PCGDP, and TFR in Table 19.8.

countries to the right. In fact, the mean is \$5,447 and the median is just \$1,302.

And we can now see that the variable TFR is distributed bimodally. This is especially evident in the frequency polygon, which has the telltale two-hump shape of a bimodal distribution. Many countries appear to have low TFR or high TFR, with fewer countries in the middle range.

Bimodal Distributions

Obviously, if we calculate the median or mean for a bimodal variable, we won't get a realistic picture of the central tendency in the data. To understand TFR better, I divided the data into cases above the median of 3.14 and cases below it. Then I calculated the means and medians for each of the two sets of data and ran the frequency polygons.

Figure 19.7a shows the frequency polygon for the cases above the median; figure 19.7b shows the frequency polygon for the cases below the median. Now it looks

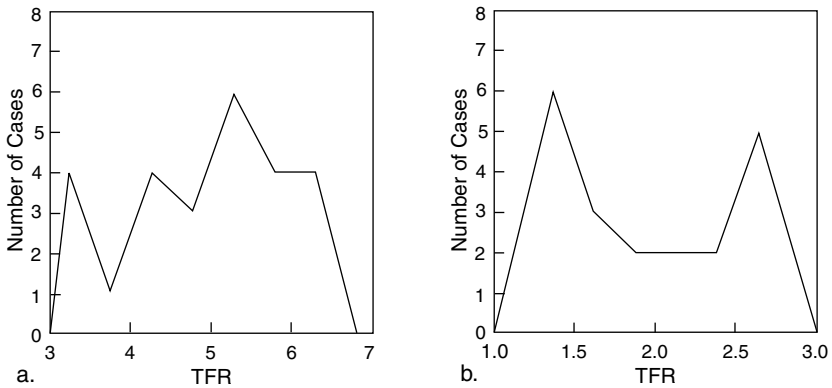


Figure 19.7. Frequency polygons for TFR (in table 19.8) above the median (figure 19.7a) and below the median (figure 19.7b).

as though we may have a multimodal distribution. This might be the result of sampling (which means we always have to be cautious about interpreting pictures of variables), but it isn't. TFR really is multimodal. I was able to find out by running the same analysis on 187 countries of the world for which we have data on this variable and comparing the shapes. Usually, this is out of the question.

Bimodal and multimodal distributions are everywhere. Figure 19.8 shows the frequency polygon for LEXFEM, female life expectancy, in table 19.8.

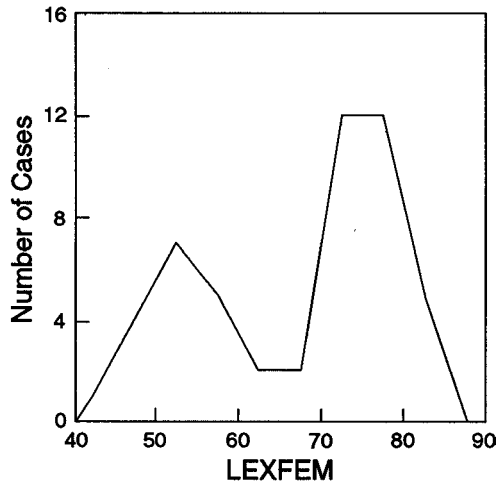


Figure 19.8. Frequency polygon for LEXFEM in Table 19.8.

This is no sampling artifact, either. The big **leptokurtic** (tall, thin) bulge on the right covers countries in which women can expect to live, on average, about 77 years. The **platykurtic** (wide, flat) bulge on the left covers countries in which women can expect to live, on average, just 55 years.

The moral is: Examine the frequency distribution for each of your variables. For interval and ordinal variables, find out if the distributions around the mean or median are symmetrical. If the distributions *are* symmetrical, then the mean is the measure of choice for central tendency. If the distributions are skewed, then the median is the measure of choice. And if the distributions are bimodal (or multimodal), then do a much closer examination and find out what's going on.

Measures of Dispersion I: Range and Interquartile Range

After central tendency and shape, the next thing we want to know about data is something about how homogeneous or heterogeneous they are—that is, something about their dispersion or variation. The concept of dispersion is easily seen in the following example. Here are the ages of two groups of five people:

Group 1:	35	35	35	35	35
Group 2:	35	75	15	15	35

Both groups have an average age of 35, but one of them obviously has a lot more variation than the other. Consider MFRATIO in table 19.8. The mean ratio of females to males in the world was 100.2 in 2000. One measure of variation in this mean across the 50 countries in our sample is the range. Inspecting the column for MFRATIO in table 19.8, we see that Latvia had the lowest ratio, with 83 males for every 100 females, and the United Arab Emirates (UAE) had the highest ratio, with 172. The range, then, is $172 - 83 = 91$. The range is a useful statistic, but it is affected strongly by extreme scores. Without the UAR, the range is 27, a drop of nearly 70%.

The **interquartile range** avoids extreme scores, either high or low. The 75th percentile for MFRATIO is 101 and the 25th percentile is 96, so the interquartile range is $101 - 96 = 5$. This tightens the range of scores, but sometimes it *is* the extreme scores that are of interest. The interquartile range of freshmen SAT scores at major universities tells you about the middle 50% of the incoming class. It doesn't tell you if the university is recruiting athletes whose SAT scores are in the bottom 25%, the middle 50%, or the top 25% of scores at those universities (see Klein 1999).

Measures of Dispersion II: Variance and the Standard Deviation

The best-known, and most-useful measure of dispersion for a sample of interval data is the **standard deviation**, usually written just s or sd . The sd is a *measure of how much, on average, the scores in a distribution deviate from the mean score*. It gives you a feel for how homogeneous or heterogeneous a population is. (We use s or sd for the standard deviation of a *sample*; we use the lower-case Greek sigma, σ , for the standard deviation of a population.)

The sd is calculated from the **variance**, written s^2 , which is the *average squared deviation from the mean* of the measures in a set of data. To find the variance in a distribution: (1) Subtract each observation from the mean of the set of observations; (2) Square the difference, thus getting rid of negative numbers; (3) Sum the differences; and (4) Divide that sum by the sample size. Here is the formula for calculating the variance:

$$s^2 = \frac{\sum (x - \bar{x})^2}{n - 1} \quad \text{Formula 19.4}$$

where s^2 is the variance, x represents the raw scores in a distribution of interval-level observations, \bar{x} is the mean of the distribution of raw scores, and n is the total number of observations.

Notice that we need to square the difference of each observation from the mean and *then* take the square root later. As we saw in calculating the mean, $\sum(x - \bar{x}) = 0$. That is, the simple sum of all the deviations from the mean is zero. Squaring each difference (each $x - \bar{x}$) gets rid of the negative numbers.

Variance describes in a single statistic how homogeneous or heterogeneous a set of data is, and, by extension, how similar or different the units of analysis described by those data are. Consider the set of scores in table 19.2 on the variable called REDUCE. These scores show people's support for the idea that "Americans are going to have to drastically reduce their consumption over the next few years." Suppose that for each level of education, you could predict the level of support for that attitudinal item about cutting back on consumption. If you could do this in 100% of all cases, then you would speak of "explaining all the variance" in the dependent variable.

I've never encountered this strength of association between two variables in the social sciences, but some things come pretty close, and in any event, the principle is what's important.

The standard deviation, s , is the square root of the variance, s^2 . The formula for the standard deviation is:

$$s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}} \quad \text{Formula 19.5}$$

Table 19.10 shows how to calculate the standard deviation for the data on female illiteracy in table 19.7. The sum of the $(x - \bar{x})^2$ is 3,896.11 and $n = 10$.

Substituting in the formula for standard deviation, we get:

$$s = \sqrt{\frac{3,896.11}{9}} = 20.81$$

If we were reporting these data, we would say that "the average percentage of adult female illiteracy is 19.49 with *sd* 20.81."

For grouped data, we take the midpoint of each interval as the raw score and use formula 19.6 to calculate the standard deviation:

$$s = \sqrt{\frac{\sum f(x - \bar{x})^2}{n - 1}} \quad \text{Formula 19.6}$$

Table 19.11 shows the procedure for calculating the standard deviation for the grouped data in table 19.4. We know from table 19.6 that $\bar{x} = 45$ for the data in table 19.4.

TABLE 19.10

How to Calculate the Standard Deviation for the Data on Female Illiteracy in Table 19.7

<i>COUNTRY</i>	<i>FEMILLIT</i> <i>x</i>	$(x - \bar{x})^2$
Armenia	2.4	$(2.4 - 19.49)^2 = 292.07$
Chad	66.0	$(66.0 - 19.49)^2 = 2,163.18$
El Salvador	23.8	$(23.8 - 19.49)^2 = 18.58$
Ghana	37.1	$(37.1 - 19.49)^2 = 310.11$
Iran	30.1	$(30.1 - 19.49)^2 = 112.57$
Latvia	0.2	$(0.2 - 19.49)^2 = 372.10$
Namibia	18.8	$(18.8 - 19.49)^2 = .48$
Panama	8.7	$(8.7 - 19.49)^2 = 116.42$
Slovenia	0.4	$(0.4 - 19.49)^2 = 364.43$
Suriname	7.4	$(7.4 - 19.49)^2 = 146.17$
		$\sum(x - \bar{x})^2 = 3,896.11$
		$s = \sqrt{3896.11/9} = 20.81$

Substituting in formula 19.6 for *sd*, we get:

$$s = \sqrt{\frac{5,800}{29}} = 14.14$$

and we report that “the mean age is 45.00, *sd* 14.14.” For comparison, the mean age and *sd* of the 30 ages (the ungrouped data) in table 19.3b are 45.033 and 15.52. Close, but not dead on.

Are these numbers describing ages and percentages of illiteracy large, or small, or about normal? There is no way to tell except by comparison across cases. By themselves, numbers like means and standard deviations simply

TABLE 19.11

Calculating the Standard Deviation for the Grouped Data in Table 19.4

<i>AGE X</i>	<i>f</i>	<i>Midpoint x</i>	$x - \bar{x}$	$(x - \bar{x})^2$	$f(x - \bar{x})^2$
20–29	6	25	$25 - 45 = -20$	400	2,400
30–39	6	35	$35 - 45 = -10$	100	600
40–49	5	45	$45 - 45 = 0$	0	0
50–59	8	55	$55 - 45 = 10$	100	800
60+	5	65	$65 - 45 = 20$	400	2,000
	$\sum x = 30$		$\sum(x - \bar{x}) = 0$		$\sum f(x - \bar{x})^2 = 5,800$

describe a set of data. But in comparative perspective, they help us produce theory; that is, they help us develop ideas about what causes things, and what those things, in turn, cause. We'll compare cases when we get to bivariate analysis in the chapter coming up next.

The Logic of Hypothesis Testing

One thing we can do, however, is test whether the *mean of a sample of data*, \bar{x} , is likely to represent the *mean of the population*, μ , from which the sample was drawn. We'll test whether the mean of FEMILLIT in table 19.7 is likely to represent the mean of the population of 50 countries. To do this, we will use the logic of hypothesis testing. This logic is used very widely—not just in the social sciences, but in all probabilistic sciences, including meteorology and genetics, for example.

The key to this logic is the statement that we can test whether the mean of the sample *is likely to represent* the mean of the population. Here's how the logic works.

1. First, we set up a **null hypothesis**, written H_0 , which states that there is no difference between the sample mean and the mean of the population from which the sample was drawn.
2. Then we set up the **research hypothesis** (also called the **alternative hypothesis**), written H_1 , which states that, in fact, the sample mean and the mean of the population from which the sample was drawn are different.
3. Next, we decide whether the research hypothesis is only about **magnitude** or is **directional**. If H_1 is only about magnitude—that is, it's **nondirectional**—then it can be stated just as it was in (2) above: The sample mean and the mean of the population from which the sample was drawn are different. Period.

If H_1 is directional, then it has to be stated differently: The sample mean is [bigger than] [smaller than] the mean of the population from which the sample was drawn.

This decision determines whether we will use a **one-tailed** or a **two-tailed** test of the null hypothesis.

To understand the concept of one-tailed and two-tailed tests, suppose you have a bell curve that represents the distribution of means from many samples of a population. Sample means are like any other variable. Each sample has a mean, and if you took thousands of samples from a population you'd get a distribution of means (or proportions). Some would be large, some small, and some exactly the same as the true mean of the population. The distribution

would be normal and form a bell curve like the one at the top of figure 19.4 and in figure 7.1.

The unlikely means (the very large ones and the very small ones) show up in the narrow area under the tails of the curve, while the likely means (the ones closer to the true mean of the population) show up in the fat, middle part. In research, the question you want to answer is whether the means of variables from one, particular sample (the one *you've* got) probably represent the tails or the middle part of the curve.

Hypothesis tests are two tailed when you are interested only in whether the magnitude of some statistic is significant (i.e., whether you would have expected that magnitude by chance). When the direction of a statistic is not important, then a two-tailed test is called for.

As we'll see in chapter 20, however, when you predict that one of two means will be higher than the other (like predicting that the mean of the final exam will be higher than the mean of the midterm exam in a class) you would use a one-tailed test. After all, you'd be asking only whether the mean was likely to fall in one tail of the normal distribution. Look at appendix C carefully. Scores significant at the .10 level for a two-tailed test are significant at the .05 level for a one-tailed test.

4. Finally, we determine the **alpha level**, written α , which is the **level of significance** for the hypothesis test. Typically, alpha is set at the .05 level or at the .01 level of significance. What this means is that if a mean or a proportion from a sample is likely to occur more than alpha—say, more than 5% of the time—then we *fail to reject the null hypothesis*.

And conversely: If the mean or a proportion of a sample is likely to occur by chance less than alpha, then we *reject the null hypothesis*. Alpha defines the **critical region** of a sampling distribution—that is, the fraction of the sampling distribution small enough to reject the null hypothesis.

In neither case do we *prove* the research hypothesis, H_1 . We either reject or fail to reject the null hypothesis. Failing to reject the null hypothesis is the best we can do, since, in a probabilistic science, we can't ever really prove any research hypothesis beyond any possibility of being wrong.

On Being Significant

By custom—and only by custom—researchers generally accept as **statistically significant** any outcome that is not likely to occur by chance more than five times in a hundred tries. This *p*-value, or **probability value**, is called the **.05 level of significance**. A *p*-value of .01 is usually considered *very* significant, and .001 is often labeled *highly* significant.

Many researchers use asterisks instead of p -values in their writing to cut down on number clutter. A single asterisk signifies a p -value of .05, a double asterisk signifies a value of .01 or less, and a triple asterisk signifies a value of .001 or less. If you read: “Men were more likely than women** to report dissatisfaction with local schoolteacher training,” you’ll know that the double asterisk means that the difference between men and women on this variable was significant at the .01 level or better.

And remember: Statistical significance is one thing, but *substantive* significance is another matter entirely. In exploratory research, you might be satisfied with a .10 level of significance. In evaluating the side effects of a medical treatment, you might demand a .001 level—or an even more stringent test of significance.

Type I and Type II Errors

There is one more piece to the logic of hypothesis testing. The choice of an alpha level lays us open to making one of two kinds of error—called, conveniently, **Type I error** and **Type II error**.

If we reject the null hypothesis when it’s really true, that’s a Type I error. If we fail to reject the null hypothesis when it is, in fact, false, that’s a Type II error.

Suppose we set alpha at .05 and make a Type I error. Our particular sample produced a mean or a proportion that happened to fall in one of the 2.5% tails of the distribution (2.5% on either side accounts for 5% of all cases), but 95% of all cases would have resulted in a mean that let us reject the null hypothesis.

Suppose the government of a small Caribbean island wants to increase tourism. To do this, somebody suggests implementing a program to teach restaurant workers to wash their hands after going to the bathroom and before returning to work. The idea is to lower the incidence of food-borne disease and instill confidence among potential tourists.

Suppose you do a pilot study and the results show that H_0 is false at the .05 level of significance. The null hypothesis is that the proposed program is useless, so this would be great news. But suppose H_0 is really true—the program really is useless—but you reject it. This Type I error sets off a flurry of activity: The Ministry of Health requires restaurant owners to shell out for the program. And for what?

The obvious way to guard against this Type I error is to raise the bar and set α at, say, .01. That way, a Type I error would be made once in a hundred tries, not five times in a hundred. But you see immediately the cost of this little ploy: It increases dramatically the probability of making a Type II error—not

rejecting H_0 when we should do exactly that. In this case, the result is that you place people at greater risk of food-borne disease.

In a probabilistic science, we are always in danger of making one or the other of these errors. Do we try to avoid one kind more than the other? It depends on what's at stake. A Type I error at the .01 level for an HIV test means that one person out of a hundred is declared HIV-free when they are really HIV-positive. How big a sample do you need to get that error level down to one out of a thousand? This is the kind of question answered by **power analysis**, which we'll look at in chapter 20.

So, What about the Mean of FEMILLIT?

As you know from chapter 7, statistics (like the mean) vary from sample to sample and how much they vary depends on: (1) the size of the sample and (2) the amount of actual variation in the population from which you take your sample. The average amount of error we make in estimating a parameter from sample statistics is called the **standard error**, or SE, of the statistic. The standard error of the mean, SEM, is the standard deviation, sd , divided by the square root of the sample size, n :

$$SEM = \frac{sd}{\sqrt{n}} \quad \text{Formula 19.7}$$

We can calculate the SEM for the sample of 10 countries on the variable FEMILLIT. From table 19.10, we know that $\bar{x} = 19.49\%$ and $sd = 20.81$ so:

$$SEM = \frac{20.81}{\sqrt{10}} = 6.58$$

Now, knowing the SEM, we can ask whether the mean of the sample of 10 cases in table 19.7 (19.49%) can be distinguished statistically from the mean of the population of countries in table 19.8.

Since this is a small sample, we can test this using Student's t distribution, which I introduced in chapter 7 (see figure 7.7). We have data on the percentage of adult female illiteracy for 45 countries in table 19.8, and the mean for *those* data is 25.51%. The formula for calculating t , when the parameter, μ , is known, is:

$$t = \frac{\bar{x} - \mu}{SEM} \quad \text{Formula 19.8}$$

So, for the sample of 10 countries on FEMILLIT,

$$t = (19.49 - 25.51)/6.58 = -0.91$$

We can test whether the value of t is statistically significant by referring to the t -table in appendix C. To use appendix C, we need to know: (1) how many degrees of freedom we have and (2) whether we want a one- or a two-tailed test.

To understand the concept of degrees of freedom, suppose I give you a jar filled with thousands of beans numbered from 1 to 9 and ask you to pick two that sum to 10. If you pick a 4 on the first draw, then you must pick a 6 on the next; if you pick a 5 on the first draw, then you must pick another 5; and so on. This is an example of one degree of freedom, because after the first draw you have no degrees of freedom left.

Suppose, instead, that I ask you to pick four beans that sum to 25. In this example, you have three degrees of freedom. No matter what you pick on the first draw, there are lots of combinations you can pick on the next three draws and still have the beans sum to 25. But if you pick a 6, a 9, and a 7 on the first *three* draws, then you must pick a 3 on the last draw. You've run out of degrees of freedom.

For a one-sample, or univariate t -test, the degrees of freedom, or df , is simply $n - 1$. For the sample of 10 representing FEMILLIT there are $10 - 1 = 9$ degrees of freedom.

Testing the Value of t

We'll use a two-tailed test for the problem here because we are only interested in whether our sample mean, 19.49, is significantly *different from*, the population mean of 25.51. Thus, the null hypothesis is that 19.49% could be found just by chance if we drew this sample of 10 countries from a population of countries where the mean is 25.51%.

Looking at the values in appendix C, we see that any t -value above 2.262 is statistically significant at the .05 level with 9 degrees of freedom for a two-tailed test. With a t -value of -0.91 (we only look at the absolute value and ignore the minus sign), we can not reject the null hypothesis. The true mean percentage of female illiteracy rates across the 50 countries of the world is statistically indistinguishable from 19.49%.

Testing the Mean of Large Samples

Another way to see this is to apply what we learned about the normal distribution in chapter 7. We know that in any normal distribution for a large popu-

lation, 68.26% of the statistics for estimating parameters will fall within one standard error of the actual parameter; 95% of the estimates will fall between the mean and 1.96 standard errors; and 99% of the estimates will fall between the mean and 2.58 standard errors.

In table 19.2, I showed you a small sample of the data from the study that Ryan, Borgatti, and I did (Bernard et al. 1999) on attitudes about environmental activism. In that study, we interviewed a random sample of 609 adults from across the United States. The mean age of respondents in our sample was 44.21, *sd* 15.75. Only 591 respondents agreed to tell us their age, so the standard error of the mean is:

$$15.75 / \sqrt{591} = 0.648$$

Since we have a large sample, we can calculate the 95% confidence limits using the *z* distribution in appendix B:

$$44.21 \pm 1.96(0.648) = 44.21 - 1.27 = 42.94 \text{ and } 44.21 + 1.27 = 45.48$$

In other words, we expect that 95% of all samples of 591 taken from the population of adults in the United States will fall between 42.94 and 45.48. As we saw in chapter 7, these numbers are the 95% **confidence limits** of the mean. As it happens, we know from the U.S. Census Bureau that the real average age of the adult (over-18) population in the United States in 1997 (when the survey was done) was 44.98.

Thus: (1) the sample statistic ($\bar{x} = 44.21\%$) and (2) the parameter ($\mu = 44.98\%$) *both* fall within the 95% confidence limits, and we *can not reject the null hypothesis* that our sample comes from a population whose average age is equal to 44.98%.

More about *z*-Scores

As we saw also in chapter 7 on sampling, every real score in a distribution has a *z*-score, also called a **standard score**. A *z*-score tells you how far, in standard deviations, a real score is from the mean of the distribution. The formula for finding a *z*-score is:

$$z = \frac{(\text{raw score} - \bar{x})}{SD} \qquad \textbf{Formula 19.9}$$

To find the *z*-scores of the data on FEMILLIT in table 19.7, subtract 19.49 (the mean) from each raw score and divide the result by 20.81 (the *sd*). Table 19.12 shows these *z*-scores.

TABLE 19.12
z-Scores of the Data on FEMILLIT in Table 19.7

<i>COUNTRY</i>	$x - \bar{x}$	<i>z-score</i>
Armenia	2.4 - 19.49 = -17.09	-0.8212
Chad	66.0 - 19.49 = 46.51	2.2349
El Salvador	23.8 - 19.49 = 4.31	0.2071
Ghana	37.1 - 19.49 = 17.61	0.8462
Iran	30.1 - 19.49 = 10.61	0.5100
Latvia	0.2 - 19.49 = -19.29	-0.9270
Namibia	18.8 - 19.49 = -0.69	-0.0332
Panama	8.7 - 19.49 = -10.79	-0.5185
Slovenia	0.4 - 19.49 = -19.09	-0.9173
Suriname	7.4 - 19.49 = -12.09	-0.5810

Why Use Standard Scores?

There are several advantages to using standard scores rather than raw scores. First of all, while raw scores are always in specialized units (percentages of people, kilos of meat, hours of time, etc.), standard scores measure the difference, in standard deviations, between a raw score and the mean of the set of scores. A *z*-score close to 0 means that the raw score was close to the average. A *z*-score that is close to plus-or-minus 1 means that the raw score was about one standard deviation from the mean, and so on.

What this means, in practice, is that when you standardize a set of scores, you create a scale that lets you make comparisons *within* chunks of your data.

For example, we see from table 19.12 that the female illiteracy rates for Namibia and Ghana are 18.8% and 37.1%, respectively. One of these raw numbers is about double the other. Similarly, the female illiteracy rate for El Salvador is almost three times that of Panama (see table 19.8).

These raw numbers tell us something, but the *z*-scores tell us more: The percentage of female illiteracy in Namibia is close to the mean for this group of countries, but the percentage for Ghana is almost 1 *sd* (.85) above the mean. The percentage for El Salvador is about .2 *sd* above the mean, and the percentage for Slovenia is almost 1 *sd* (-.92) below the mean.

A second advantage of standard scores over raw measurements is that standard scores are independent of the units in which the original measurements are made. This means that you can compare the relative position of cases across different variables.

Medical anthropologists measure variables called “weight-for-length” and “length-for-age” in the study of nutritional status of infants across cultures. Linda Hodge and Darna Dufour (1991) studied the growth and development

of Shipibo Indian children in Peru. They weighed and measured 149 infants, from newborns to 36 months in age.

By converting all measurements for height and weight to z -scores, they were able to compare their measurements of the Shipibo babies against standards set by the World Health Organization (Frisancho 1990) for healthy babies. The result: By the time Shipibo children are 12 months old, 77% of boys and 42% of girls have z -scores of -2 or more on length-for-age. In other words, by a year old, Shipibo babies are more than two standard deviations under the mean for babies who, by this measure, are healthy.

By contrast, only around 10% of Shipibo babies (both sexes) have z -scores of -2 or worse on weight-for-length. By a year, then, most Shipibo babies are clinically stunted but they are not clinically wasted. This does not mean that Shipibo babies are small but healthy. Infant mortality is as high as 50% in some villages, and the z -scores on all three measures are similar to scores found in many developing countries where children suffer from malnutrition.

z -scores do have one disadvantage: They can be downright unintuitive. Imagine trying to explain to people who have no background in statistics why you are proud of having scored a 1.96 on the SAT. Getting a z -score of 1.96 means that you scored almost two standard deviations above the average and that only 2.5% of all the people who took the test scored higher than you did. If you got a z -score of $-.50$ on the SAT, that would mean that about a third of all test takers scored lower than you did. That's not too bad, but try explaining why you're not upset about getting a *minus* score on *any* test.

This is why T -scores (with a capital T —not to be confused with Student's t) were invented. The mean of a set of z -scores is always 0 and its standard deviation is always 1. T -scores are **linear transformations** of z -scores. For the SAT and GRE, the mean is set at 500 and the standard deviation is set at 100. A score of 400 on these tests, then, is one standard deviation below the mean; a score of 740 is 2.4 standard deviations above the mean (Friedenberg 1995:85).

The Univariate Chi-Square Test

Finally, chi-square (often written χ^2) is an important part of univariate analysis. It is a test of whether the distribution of a series of counts is likely to be a chance event. The formula for χ^2 is:

$$\chi^2 = \sum \frac{(O - E)^2}{E} \quad \text{Formula 19.10}$$

where O represents the observed number of cases and E represents the number of cases you'd expect, *ceteris paribus*, or "all other things being equal."

Suppose that among 14 families there are a total of 42 children. If children were distributed equally among the 14 families, we'd expect each family to have three of them. Table 19.13 shows what we would expect and what we might find in an actual set of data. The χ^2 value for this distribution is 30.65.

Finding the Significance of χ^2

To determine whether this value of χ^2 is statistically significant, first calculate the degrees of freedom (abbreviated *df*) for the problem. For a univariate table:

$$df = \text{the number of cells, minus one}$$

or $14 - 1 = 13$ in this case.

Next, go to appendix D, which is the distribution for χ^2 , and read down the left-hand margin to 13 degrees of freedom and across to find the *critical value* of χ^2 for any given level of significance. The levels of significance are listed across the top of the table.

The greater the significance of a χ^2 value, the less likely it is that the distribution you are testing is the result of chance.

In exploratory research, you might be satisfied with a .10 level of significance. In evaluating the side effects of a medical treatment you might demand a .001 level—or even more. Considering the χ^2 value for the problem in table 19.13, the results look pretty significant. With 13 degrees of freedom, a χ^2 value of 22.362 is significant at the .05 level; a χ^2 value of 27.688 is significant at the .01 level; and a χ^2 value of 34.528 is significant at the .001 level. With a χ^2 of 30.65, we can say that the distribution of the number of children across the 14 families is statistically significant at better than the .01 level, but not at the .001 level.

Statistical significance here means only that the distribution of number of children for these 14 families is not likely to be a chance event. Perhaps half the families happen to be at the end of their fertility careers, while half are just starting. Perhaps half the families are members of a high-fertility ethnic group and half are not. The *substantive* significance of these data requires interpretation, based on your knowledge of what's going on, on the ground.

Univariate numerical analysis—frequencies, means, distributions, and so on—and univariate graphical analysis—histograms, box plots, frequency polygons, and so on—tell us a lot. Begin all analysis this way and let all your data and your experience guide you in their interpretation. It is not always possible, however, to simply scan your data and use univariate, descriptive statistics to understand the subtle relations that they harbor. That will require more complex techniques, which we'll take up in the next two chapters.

TABLE 19.13
Chi-Square for a Univariate Distribution

Expected number of children per family															
Family #	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total
	3	3	3	3	3	3	3	3	3	3	3	3	3	3	42
Observed number of children per family															
Family #	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total
	0	0	5	5	5	6	6	0	0	3	1	0	6	5	42
(Observed - Expected) ²															
9	9	9	4	4	9	9	9	9	9	0	4	9	9	4	
(Observed - Expected) ²															
3	3	1.33	1.33	1.33	3	3	3	3	3	0	1.33	3	3	1.33	
$\chi^2 = \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}} = 3 + 3 + 1.33 + 1.33 + \dots + 1.33 = 30.65$															

20



Bivariate Analysis: Testing Relations

This chapter is about describing relations between pairs of variables—**covariations**—and testing the significance of those relations.

The *qualitative* concept of covariation creeps into ordinary conversation all the time: “If kids weren’t exposed to so much TV violence, there would be less crime.” Ethnographers also use the concept of covariation in statements like: “Most women said they really wanted fewer pregnancies, but claimed that this wasn’t possible so long as the men required them to produce at least two fully grown sons to work the land.” Here, the number of pregnancies is said to covary with the number of sons husbands say they need for agricultural labor.

The concept of *statistical* covariation, however, is more precise. There are two primary and two secondary things we want to know about a statistical relation between two variables:

The primary questions are these:

1. How big is it? That is, how much better could we predict the score of a dependent variable in our sample if we knew the score of some independent variable? Correlation coefficients answer this question.
2. Is the covariation due to chance, or is it statistically significant? In other words, is it likely to exist in the overall population to which we want to generalize? Statistical tests answer this question.

For many problems, we also want to know:

3. What is its direction? Is it positive or negative?
4. What is its shape? Is it linear or nonlinear?

Answers to these questions about qualities of a relationship come best from looking at graphs.

Testing for **statistical significance** is a mechanical affair—you look up, in a table, whether a statistic showing covariation between two variables is, or is not statistically significant. I'll discuss how to do this for several of the commonly used statistics that I introduce below. Statistical significance, however, does not necessarily mean substantive or theoretical importance. Interpreting the substantive and theoretical importance of statistical significance is anything but mechanical. It requires thinking. And that's *your* job.

The *t*-Test: Comparing Two Means

We begin our exploration of bivariate analysis with the two-sample *t*-test. In chapter 19, we saw how to use the one-sample *t*-test to evaluate the probability that the mean of a sample reflects the mean of the population from which the sample was drawn. The two-sample *t*-test evaluates whether the means of two independent groups differ on some variable. Table 20.1 shows some data that Penn Handwerker collected in 1978 from American and Liberian college students on how many children those students wanted.

Table 20.2 shows the relevant statistics for the data in table 20.1. (I generated these with SYSTAT®, but you can use any statistics package.)

There are 43 American students and 41 Liberian students. The Americans wanted, on average, 2.047 children, *sd* 1.396, SEM 0.213. The Liberians wanted, on average, 4.951 children, *sd* 2.387, SEM 0.373.

The null hypothesis, H_0 , is that these two means, 2.047 and 4.951, come from random samples of the *same* population—that there is no difference, except for sampling error, between the two means. Stated another way, these two means come from random samples of two populations with identical averages. The research hypothesis, H_1 , is that these two means, 2.047 and 4.951, come from random samples of truly different populations.

The formula for calculating *t* for two independent samples is:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2(1/n_1 + 1/n_2)}} \quad \text{Formula 20.1}$$

That is, *t* is the difference between the means of the samples, divided by the fraction of the standard deviation, σ , of the total population, that comes from each of the two separate populations from which the samples were drawn. (Remember, we use Roman letters, like *s*, for sample statistics, and Greek letters, like σ , for parameters.) Since the standard deviation is the square root of the variance, we need to know the variance, σ^2 , of the **parent population**.

TABLE 20.1

Number of Children Wanted (CW) by College Students in the United States and in Liberia

#	CW	Region	Sex	#	CW	Region	Sex	#	CW	Region	Sex
1	1	USA	M	29	0	USA	F	57	12	WA	M
2	3	USA	M	30	0	USA	F	58	12	WA	M
3	3	USA	M	31	0	USA	F	59	8	WA	M
4	3	USA	M	32	0	USA	F	60	6	WA	M
5	3	USA	M	33	1	USA	F	61	4	WA	M
6	3	USA	M	34	1	USA	F	62	4	WA	M
7	2	USA	M	35	2	USA	F	63	2	WA	M
8	2	USA	M	36	2	USA	F	64	3	WA	M
9	2	USA	M	37	2	USA	F	65	3	WA	M
10	2	USA	M	38	2	USA	F	66	4	WA	M
11	2	USA	M	39	2	USA	F	67	4	WA	M
12	2	USA	M	40	2	USA	F	68	4	WA	M
13	1	USA	M	41	2	USA	F	69	4	WA	M
14	6	USA	M	42	2	USA	F	70	3	WA	F
15	1	USA	M	43	2	USA	F	71	3	WA	F
16	1	USA	M	44	6	WA	M	72	3	WA	F
17	4	USA	M	45	4	WA	M	73	3	WA	F
18	0	USA	M	46	4	WA	M	74	7	WA	F
19	5	USA	F	47	4	WA	M	75	2	WA	F
20	4	USA	F	48	5	WA	M	76	4	WA	F
21	4	USA	F	49	5	WA	M	77	6	WA	F
22	4	USA	F	50	5	WA	M	78	4	WA	F
23	3	USA	F	51	5	WA	M	79	4	WA	F
24	2	USA	F	52	5	WA	M	80	4	WA	F
25	3	USA	F	53	7	WA	M	81	4	WA	F
26	0	USA	F	54	12	WA	M	82	4	WA	F
27	2	USA	F	55	6	WA	M	83	4	WA	F
28	0	USA	F	56	6	WA	M	84	4	WA	F

SOURCE: *Data Analysis with MYSTAT* by H. R. Bernard and W. P. Handwerker, 1995. McGraw-Hill. Reproduced with permission of The McGraw-Hill Companies.

The parent population is the general population from which the two samples were pulled. Our best guess at σ^2 is to pool the standard deviations from the two samples:

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \quad \text{Formula 20.2}$$

which is very messy, but just a lot of arithmetic. For the data on the two groups of students, the **pooled variance** is:

TABLE 20.2
Descriptive Statistics for Data in Table 20.1

	<i>CW-USA</i>	<i>CW-Liberia</i>
N of cases	43	41
Minimum	0	2
Maximum	6	12
Mean	2.047	4.951
95% CI Upper	2.476	5.705
95% CI Lower	1.617	4.198
Standard Error	0.213	0.373
Standard Deviation	1.396	2.387

$$s^2 = \frac{(43 - 1)1.396^2 + (41 - 1)2.387^2}{43 + 41 - 2} = \frac{81.85 + 227.91}{82} = 3.778$$

Now we can solve for *t*:

$$t = \frac{2.047 - 4.951}{\sqrt{3.778(.0477)}} = \frac{-2.904}{\sqrt{.180}} = \frac{-2.904}{.4243} = 6.844$$

Testing the Value of t

We can evaluate the statistical significance of *t* using appendix C. Recall from chapter 19 that we need to calculate the degrees of freedom and decide whether we want a one- or a two-tailed test to find the critical region for rejecting the null hypothesis. For a two-sample *t*-test, the degrees of freedom equals:

$$(n_1 + n_2) - 2$$

so there are 43 + 41 - 2 = 82 degrees of freedom in this particular problem.

If you test the possibility that one mean will be higher than another, then you need a one-tailed test. After all, you're only asking whether the mean is likely to fall in one tail of the *t*-distribution (see figure 7.7). With a one-tailed test, a finding of no difference (the null hypothesis) is equivalent to finding that you predicted the wrong mean to be the one that was higher. If you want to test only whether the two means are different (and not that one will be higher than the other), then you need a two-tailed test. Notice in appendix C that scores significant at the .10 level for a two-tailed test are significant at the .05 level for a one-tailed test; scores significant at the .05 level for a two-tailed test are significant at the .025 level for a one-tailed test, and so on.

We'll use a two-tailed test for the problem here because we are only inter-

ested in the magnitude of the difference between the means, not its direction or sign (plus or minus). We are only interested here, then, in the **absolute value** of t , 6.844.

Looking at the values in appendix C, we see that any t -value above 3.291 is significant for a two-tailed test at the .001 level. Assuming that our samples represent the populations of American students and Liberian students, we'd expect the observed difference in the means of how many children they want to occur by chance less than once every thousand times we run this survey.

Analysis of Variance

A t -test measures the difference between two means. Analysis of variance, or ANOVA, is a technique that applies to a set of two or more means. Table 20.3 shows the scores for four groups of states in the United States on the percentage of teenage births in 1996. Each group of states has a mean percentage of teenage births. The question we want to answer is: Are the differences among the means significantly different, at say, the .05 level?

To answer this question we will calculate the ratio of the **between-group variance**, s_{between}^2 , to the **within-group variance**, s_{within}^2 . The between-group and within-group variances sum to the **total variance**, s^2 . We already know, from chapter 19, the formula for the total variance. Here it is again:

$$s^2 = \frac{\sum (x - \bar{x})^2}{n - 1} \quad \text{Formula 20.3}$$

where x represents the raw scores in a distribution of interval-level observations, \bar{x} is the mean of the distribution of raw scores, and n is the total number of observations. So, to find the variance in a distribution: (1) subtract each observation from the mean of the set of observations; (2) square the difference (thus getting rid of negative numbers); (3) sum the squared differences (this is called the **sum of the squares**); and (4) divide that sum by the sample size, minus 1.

The variance is calculated from the sum of the squares, $\sum (x - \bar{x})^2$. Here is the formula for calculating the sum of the squares directly from the data in table 20.3:

$$\sum x^2 - \frac{(\sum x)^2}{n} \quad \text{Formula 20.4}$$

TABLE 20.3
Percentage of Teenage Births in the 50 U.S. States

Northeast		Midwest		South		West					
x_1	x_1^2	x_2	x_2^2	x_3	x_3^2	x_4	x_4^2				
NY	9.2	84.64	KS	13.1	171.61	FL	13.4	179.56	NM	17.9	320.41
VT	8.9	79.21	SD	11.5	132.25	AR	19.8	392.04	OR	13.2	174.24
NJ	7.7	59.29	WI	10.6	112.36	KY	17.0	289.00	ID	13.5	182.25
PA	10.6	112.35	IA	11.0	121.00	LA	18.9	357.21	HI	10.3	106.09
MA	7.3	53.29	IN	14.5	210.25	MD	10.3	106.09	NV	13.3	176.89
NH	7.4	54.76	OH	13.3	176.89	MS	21.3	453.69	CO	11.9	141.61
ME	9.7	94.09	ND	9.6	92.16	AL	18.3	334.89	UT	10.6	112.36
CT	8.2	67.24	MI	12.2	148.84	DE	13.7	187.69	CA	12.0	144.00
RI	10.3	106.09	MN	8.5	72.25	GA	15.9	252.81	WY	14.4	207.36
			MO	14.4	207.36	OK	17.2	295.84	WA	11.3	127.69
			NE	10.6	112.36	VA	11.0	121.00	AZ	15.0	225.00
			IL	12.7	161.29	NC	15.0	225.00	AK	11.2	125.44
						WV	16.8	282.24	MT	12.5	156.25
						TN	16.8	282.24			
						SC	16.8	282.24			
						TX	16.2	262.44			
$\sum x_1 = 79.3$		$\sum x_2 = 142.0$		$\sum x_3 = 258.4$		$\sum x_4 = 167.1$					
$\sum x_1^2 = 710.97$		$\sum x_2^2 = 1,718.62$		$\sum x_3^2 = 4,303.98$		$\sum x_4^2 = 2,199.59$					
$n = 9$		$n = 12$		$n = 16$		$n = 13$					
$\bar{x}_1 = 8.811$		$\bar{x}_2 = 11.833$		$\bar{x}_3 = 16.150$		$\bar{x}_4 = 12.854$					
$s_1^2 = 1.531$		$s_2^2 = 3.481$		$s_3^2 = 8.721$		$s_4^2 = 4.309$					

SOURCE: *Statistical Abstracts of the United States*, 1998, Table 103.

At the bottom of table 20.3, I've calculated the four separate, or within-group, variances, using formula 20.3 above. Applying formula 20.4, the total sums of squares for the data in table 20.3 is:

$$\begin{aligned}
 &710.97 + 1,718.62 + 4,303.98 + 2,199.59 - \frac{(79.3 + 142 + 258.5 + 167)^2}{50} \\
 &= 8,993.16 - \frac{418,350.24}{50} = 8,933.16 - 8,367.005 = 556.16
 \end{aligned}$$

The formula for calculating the sum of the squares between groups is:

$$\sum \frac{\sum x_{1,2,\dots,n}}{n_{1,2,\dots,n}} - \frac{\left(\sum x_{total}\right)^2}{n_{total}} \quad \text{Formula 20.5}$$

Thus, the between-group sum of the squares is:

$$(79.3^2/9) + (142^2/12) + (258.4^2/16) + (167.1^2/13) - 8,367.005 \\ = 698.72 + 1,680.33 + 4,173.16 + 2,147.88 - 8,367.005 = 333.09$$

And since:

$$\begin{aligned} \text{within-group sum of squares} &= \\ \text{total sum of squares} - \text{between-group sum of squares} \end{aligned}$$

the within-group sums of square is:

$$56.16 - 333.09 = 223.07$$

We can now calculate $s_{between}^2$, the between-group variance, and s_{within}^2 , the within-group variance.

$$s_{between}^2 = \frac{\text{total between-group sums of squares}}{\text{degrees of freedom between groups}}$$

and

$$s_{within}^2 = \frac{\text{total within-group sums of squares}}{\text{degrees of freedom within groups}}$$

We compute the degrees of freedom between groups and the degrees of freedom within groups as follows:

$$\begin{aligned} df \text{ between groups} &= \text{number of groups} - 1 \\ df \text{ within groups} &= n - \text{number of groups} \end{aligned}$$

So, for the data in table 20.3, the df between groups is $4 - 1 = 3$, and the df within groups is $50 - 4 = 46$. Then:

$$\text{the between-group variance} = 333.09/3 = 111.03$$

and

$$\text{the within-group variance} = 223.07/46 = 4.85$$

We now have all the information we need to calculate the F -ratio, which is:

$$\frac{s_{between}^2}{s_{within}^2}$$

(The F statistic was named for Sir Ronald Fisher, who developed the idea for the ratio of the between-group and the within-group variances as a general method for comparing the relative size of means across many groups.) We calculate the ratio of the variances as follows:

$$F = 111.03/4.85 = 22.89$$

If you run the ANOVA on a computer, using SYSTAT® or SPSS®, etc., you'll get a slightly different answer: 23.307. The difference (0.42) is due to rounding in all the calculations we just did. Computer programs may hold on to 12 decimal places all the way through before returning an answer. In doing these calculations by hand, I've rounded each step of the way to just two decimal places. Rounding error doesn't affect the results of the ANOVA calculations in this particular example because the value of the F statistic is so big. But when the value of the F statistic is below 5, then a difference of .42 either way can lead to serious errors of interpretation.

The moral is: Learn how to do these calculations by hand, once. Then, use a computer program to do the drudge work for you, once you know what you're doing.

Well, is 22.89 a statistically significant number? To find out, we go to appendix E, which shows the values of F for the .05 and the .01 level of significance. The values for the between-group df are shown across the top of appendix E, and the values for the within-group df are shown along the left side. Looking across the top of the table, we find the column for the between-group df . We come down that column to the value of the within-group df . In other words, we look down the column labeled 3 and come down to the row for 46.

Since there is no row for exactly 46 degrees of freedom for the within-group value, we use the nearest value, which is for 40 df . We see that any F value greater than 2.84 is statistically significant at the .01 (that is, the 1%) level. The F value we got for the data in table 20.3 was a colossal 22.89. As it turns out, this is statistically significant beyond the .001 level.

Examples of ANOVA

Camilla Harshbarger (1986) tested the productivity—measured in average-bushels-per-hectare—of 44 coffee farmers in Costa Rica, as a function of which credit bank they used or if they used no credit at all. This is the sort of problem that calls for ANOVA because Harshbarger had four different means—one for each of the three credit sources and one for farmers who chose not to use credit. The ANOVA showed that there was no significant dif-

ference in productivity in Harshbarger's sample, no matter where they got credit, or even if they didn't use credit.

Carole Jenkins (1981) surveyed 750 children in Belize for protein-calorie malnutrition (PCM). She had four means—one for each of the four ethnic groups (Creole, Mestizo, Black Carib, and Maya) in her research area. An analysis of variance showed that there was a very strong relationship between ethnic group and the likelihood of suffering from childhood PCM.

Sokolovsky et al. (1978) compared the average number of first-order relations and the average number of multiplex relations among three groups of psychiatric patients who were released to live in a hotel in midtown New York City. (First-order relations are primary relations with others; multiplex relations contain more than one kind of content, such as relations based on visiting and on borrowing money from, for example.)

One group of patients had a history of schizophrenia with residual symptoms; a second group had a history of schizophrenia without residual symptoms; and the third group had no psychotic history. An ANOVA showed clearly that the average network size (both first-order and multiplex networks) was different among the three groups. From these data (and from field observation and in-depth interviews), Sokolovsky was able to draw strong conclusions about the ability of members of the three groups to cope with deinstitutionalization.

Whenever you observe three or more groups (age cohorts, members of different cultures or ethnic groups, people from different communities) and count anything (e.g., some behavior over a specific period of time, or the number of particular kinds of contacts they make, or the number of kilograms of fish they catch), then ANOVA is the analytic method of choice. If you are interested in the causes of morbidity, for example, you could collect data on the number of sick days among people in various social groups over a given period of time. Other dependent variables in which anthropologists are interested, and which are amenable to ANOVA, are things like blood pressure, number of minutes per day spent in various activities, number of grams of nutrients consumed per day, and scores on tests of knowledge about various cultural domains (plants, animals, diseases), to name just a few.

When there is one dependent variable (such as a test score) and one independent variable (a single intervention like the reading program), then no matter how many groups or tests are involved, a **one-way ANOVA** is needed. If more than one independent variable is involved (say, several competing new housing programs, and several socioeconomic backgrounds), and there is a single dependent variable (a reading test score), then **multiple-way ANOVA**, or **MANOVA**, is called for. When two or more dependent variables are corre-

lated with one another, then **analysis of covariance**, or ANCOVA techniques are used.

Direction and Shape of Covariations

As you can tell from my discussion of box plots and frequency polygons and such in chapter 19, I like to look at things like shape and direction—qualitative things—in connection with numerical results.

For example, the amount of cholesterol you have in your blood and the probability that you will die of a heart attack at any given age are **positive covariants**: The *more* cholesterol, the *higher* the probability.

By contrast, *the more education* you have, *the lower the probability that you smoke cigarettes*. Education and the probability of smoking cigarettes are **negative covariants**.

Some shapes and directions of bivariate relations are shown in the four **scatterplots** (also called **scattergrams**) of figure 20.1. Figure 20.1a is a plot of the number of violent crimes and the number of physicians per 100,000 people in each of the 50 states in the United States. The dots are scattered haphazardly, and it's pretty obvious that there is *no relation* between these two variables.

Linear and Nonlinear Relations

Figure 20.1b shows the correlation between LEXFEM (life expectancy for women) and INFMORT (infant mortality) in table 19.8. This is a clear case of a **linear and negative relation** between two variables: The higher the score on one variable, the lower the score tends to be on the other. We see in figure 20.1b that in countries where women's life expectancy is high, infant mortality tends to be low, and vice versa.

By convention, the values of dependent variables are plotted on the Y axis, or **abscissa** of a scatterplot, and the values of independent variables are plotted on the X axis, or **ordinate**. Notice that in this case, it is not really clear which is the dependent and which is the independent variable. The relation between the two variables in figure 20.1b is strong and clear, but recall from chapter 2 that we need much more than a strong correlation to understand cause and effect.

Figure 20.1c shows the positive correlation between TFR and FEMILLIT in table 19.8. This is a clear case of a **linear and positive relation**: The higher the score on one variable, the higher the score tends to be on the other. So, the

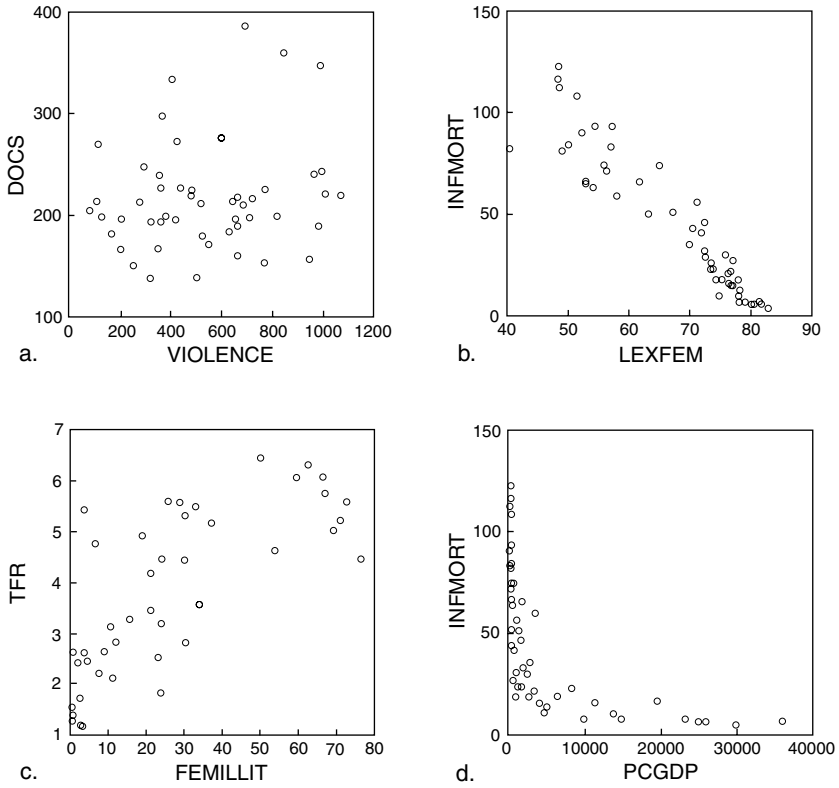


Figure 20.1. Scatterplots of bivariate relations.

higher the percentage of female illiteracy, the higher the total fertility rate across the sample of countries.

Figure 20.1d is a plot of INFMORT and PCGDP in table 19.8. The dots in figure 20.1d are not scattered around randomly—*something* is clearly going on between these two variables. As the per capita GDP rises, the rate of infant mortality falls, but the *shape* of the relation is **nonlinear**.

Nonlinear relations are very common in the real world. The relation between age and the number of people one knows has a peaked shape, like an upside-down V. Early in life, the number of friends, kin, and acquaintances is small, but that number grows as you get older. This relation is linear and positive. The longer you live, the more people you get to know . . . up to a point. If you live long enough, a lot of the people you know start dying, and your

network shrinks. There is a strong, negative relation between age and number of people in your network after age 70.

Cross-Tabs of Nominal Variables

Now that we've looked at shape and direction, let's get down to actually looking for relations between variables. We begin with nominal variables and move on to ordinal and then interval variables.

Table 20.4 shows two examples of **cross-tabs, or cross-tabulation tables** of two nominal variables. Table 20.4a is a 2×2 (read: two-by-two) cross-

TABLE 20.4a

A 2×2 Table of Household Type by Race in the United States in 1997 (in Thousands)

	<i>White</i>	<i>African American</i>	<i>Row Total</i>
Two parents	21,914 (77.6%)	1,974 (40.4%)	23,888
Single parent	6,322 (22.4%)	2,913 (59.6%)	9,235
Column total	28,236 (100%)	4,887 (100%)	33,123

SOURCE: *Statistical Abstracts of the United States, 1998, Table 79.*

tabulation of family types (two parents with children and single parents) for black and white households in the United States in 1997.

A 2×2 table is also called a **four-fold table**. Any table comparing data on two variables is called a **bivariate table**, but not all bivariate tables are 2×2 , since variables can take more than just two values. Table 20.4b shows a 3×2 (three-by-two) table. This time, family type is broken into three categories instead of just two: two parents with children, single female parents with children, and single male parents with children.

Conventions for Displaying Bivariate Tables

There are some easy-to-follow conventions for displaying tables like these. Table 20.4a is a complete table. The numbers at the margins—down the right-hand column and along the bottom row—are called, unsurprisingly, the **marginals**. The sum of the marginals down the right-hand side and the sum of the marginals across the bottom are identical, and the number in the lower-right-hand corner (33,123) is the total frequency of elements in the table.

TABLE 20.4bA 3×2 Table of Household Type by Race in the United States in 1997 (in Thousands)

	<i>White</i>	<i>African American</i>	<i>Row Total</i>
Two parents	21,914 (77.6%)	1,974 (40.4%)	23,888
Single female parent	4,997 (17.7%)	2,594 (53.1%)	7,591
Single male parent	1,325 (4.7%)	319 (6.5%)	1,644
Column total	28,236 (100%)	4,887 (100%)	33,123

SOURCE: *Statistical Abstracts of the United States*, 1998, Table 79.

I prefer to keep tables uncluttered and to show only the percentages in each cell and the n for each column. You get a better understanding of what's going on from percentages than from raw numbers in a table like this, and the interested reader can always calculate the n for each cell.

In bivariate tables, no matter what size (2×2 , 3×2 , or larger tables), the convention is to put the dependent variables in the rows and the independent variable in the columns. Then there's an easy rule to follow in reading a table: *percentage down the columns and interpret across the rows*. There are, of course, exceptions—they are, after all, conventions and not laws. When the independent variable has too many categories to fit on a narrow page, it makes sense to show the independent variables in the rows.

Percentaging down table 20.4a, we see that 77.6% of white households with children had two parents in 1997, and 22.4% of those households were headed by single parents. In black households with children, the percentages are 40.4% and 59.6%, respectively. *Interpreting across*, we see that 77.6% of white households with children had two parents in 1997 compared to 40.4% for black households. Among single-parent households, 22.4% were white and 59.6% were black.

Interpreting the numbers in a cross-tab forces you to think about explanations. The probability of a child having two parents was much higher for white children in 1996 than it was for black children—nearly two times higher, in fact.

What's going on? As I explained in chapter 2, association between two variables does not, by itself, imply cause—no matter how strong the association. The dependent variable in table 20.4 is obviously family type. Nobody's skin

color (which is, at bottom, what the so-called race variable is about in the United States) depends on whether they are a member of a two-parent or a one-parent family.

And clearly—and I mean absolutely, positively, no-fooling, clearly—being black did not cause anyone, not one single person, to be part of a single-parent household. Being a black man in the United States, however, means a high probability of attending poorly funded schools, and poor schooling produces severe disadvantage in the labor market. When men in the United States don't provide financial support, poor women—who have even worse prospects than men do for finding stable, well-paying jobs—are likely to turn to welfare.

Historically, welfare systems have punished women who have a live-in husband by lowering the women's allotments. Some women responded by maintaining single-parent households and some fraction of African American, single-parent families was, in 1997, caused by this sequence of events. There are, then, several **intervening** and **antecedent variables** that link being counted as black by the U.S. Census and being part of a single-parent household. (See figure 2.4 about intervening and antecedent variables.)

Some of the most interesting puzzles about human life involve understanding the role of intervening and antecedent variables—and figuring out which is which. Middle-aged men who drink at least six cups of coffee a day are more likely to have a heart attack than are men who don't drink coffee at all. Men who drink a lot of coffee, however, consume more alcohol, more saturated fats, and more cholesterol than men who don't drink coffee. The heavy coffee drinkers are less likely to exercise, more likely to smoke, and more likely to be impatient, aggravated people (the famous Type A personality). The jury is still out on how all these factors are related, but lots of researchers are trying to disentangle this problem (Puccio et al. 1990; Ketterer and Maercklein 1991; Schwarz et al. 1994; James 1997). (Women's chances of coronary disease are unaffected by drinking lots of coffee [Willett et al. 1996].)

Here's another one. Interethnic marriage often increases when tribal peoples move from rural villages to cities. One effect of this is that unilineal systems are under pressure to become bilateral over time (see Feldman 1994; Clignet 1966). Is it just the lack of prospective mates from one's own ethnic group that causes this? Or is there something inherently unstable about unilineal kinship systems under urban conditions?

Lambda and the PRE Principle

Table 20.5 is a hypothetical 2×2 table showing the breakdown, by gender, of adult monolingual Indians and adult bilingual Indian/Spanish speakers in a

TABLE 20.5
Monolingual and Bilingual Speakers, by Gender, in a Mexican Village, 1962

	<i>Men</i>	<i>Women</i>	<i>Row totals</i>
Bilingual	61 (82%)	24 (36%)	85
Monolingual	13 (18%)	42 (64%)	55
Column totals	74	66	140

old error = 55
new error = 13 + 24 = 37
 $\lambda = \frac{55 - 37}{55} = .33$

Mexican village in 1962. I've included the marginals and the n 's in the cells to make it easier to do the calculation here.

Reading across table 20.5, we see that 82% of the men were bilingual, compared to 36% of the women. Clearly, gender is related to whether someone is a bilingual Indian/Spanish speaker or monolingual in the Indian language only.

Suppose that for the 140 persons in table 20.5 you were asked to guess whether they were bilingual or monolingual, but you didn't know their gender. Since the mode for the dependent variable in this table is "bilingual" (85 bilinguals compared to 55 monolinguals), you should guess that everybody is bilingual. If you did that, you'd make 55 mistakes out of the 140 choices, for an error rate of 55/140, or 39%. We'll call this the **old error**.

Suppose, though, that you have all the data in table 20.5—you know the mode for gender as well as for bilingual status. Your best guess now would be that every man is bilingual and every woman is monolingual. You'd still make some mistakes, but fewer than if you just guessed that everyone is bilingual.

How many fewer? When you guess that every man is bilingual, you make exactly 13 mistakes, and when you guess that every woman is monolingual, you make 24 mistakes, for a total of 37 out of 140 or $37/140 = 26\%$. This is the **new error**. The difference between the old error (39%) and the new error (26%), divided by the old error is the **proportionate reduction of error**, or **PRE**. Thus,

$$\text{PRE} = \frac{55 - 37}{55} = .33$$

This PRE measure of association for nominal variables is called **lambda**, written either λ or L . Like all PRE measures of association, lambda has the nice quality of being intuitively and directly interpretable. A lambda of .33 means that if you know the scores on an independent variable, you can guess the scores on the dependent variable 33% more of the time than if you didn't know anything about the independent variable.

Chi-Square

While lambda demonstrates the intuitively compelling PRE principle, there are problems with it. There is no way to test whether any value of lambda shows a particularly strong or weak relationship between variables; it can take different values depending on whether you set up the dependent variable in the rows or the columns; and it can be very low, even when there is an intuitively clear association between nominal variables. Lambda for table 20.4a, for example, is just .10—that is, if you guess that all white families with children in the United States have two parents and that all black families have one parent, you make 10% fewer errors than if you guess that all families have two parents.

With bivariate data on nominal variables, many researchers use χ^2 (chi-square). Chi-square tells you whether or not a relation exists between or among variables and it tells you the probability that a relation is the result of chance. But it is *not* a PRE measure of correlation, so it doesn't tell you the *strength* of association among variables.

The principal use of χ^2 , then, is for testing the null hypothesis that there is no relation between two nominal variables. If, after a really good faith effort, we fail to accept the null hypothesis, we can reject it. Using this approach, we never prove anything using statistical tests like χ^2 . We just fail to disprove things. As it turns out, that's quite a lot.

Calculating χ^2

The formula for calculating χ^2 for a bivariate table is the same as the one we saw in chapter 19 for a univariate distribution. Here it is again:

$$\chi^2 = \sum \frac{(O - E)^2}{E} \quad \text{Formula 20.6}$$

where O represents the observed number of cases in a particular cell of a bivariate table, and E represents the number of cases you'd expect for that cell *if there were no relation* between the variables in that cell.

For each cell in a bivariate table, simply subtract the expected frequency from the observed and square the difference. Then divide by the expected frequency and sum the calculations for all the cells. Clearly, if all the observed frequencies equal all the expected frequencies, then χ^2 will be zero; that is, there will be no relation between the variables.

While χ^2 can be zero, it can never have a negative value. The more the O 's differ from the E 's (i.e., something nonrandom is going on), the bigger χ^2 gets.

Calculating the Expected Frequencies for χ^2 in Bivariate Tables

Chi-square is calculated on raw frequencies, not on percentages in tables. The expected frequencies are calculated *for each cell* with the formula:

$$F_e = \frac{(R_i)(C_j)}{n} \quad \textbf{Formula 20.7}$$

where F_e is the expected frequency for a particular cell in a table; (R_i) is the frequency total for the row in which that cell is located; (C_j) is the frequency total for the column in which that cell is located; and n is the total sample size (the lower-right-hand marginal).

The test for χ^2 can be applied to any size bivariate table. Table 20.6 shows a hypothetical census of observed adherents, in four Native American tribes, of three competing religions, and the expected number of adherents to each religion. Reading across the top of the table, in tribe #1, there are 150 Catholics, 104 Protestants, and 86 members of the Native American Church. For Tribe #1, we expect:

$$\frac{(340)(590)}{2,011} = 99.75$$

Catholics (the cell in the upper-left-hand corner of table 20.6). We expect 119.03 Protestants and 121.22 members of the Native American Church for Tribe #1, and so on.

Chi-square for this table is a whopping 162.08. To determine the number of degrees of freedom for a bivariate χ^2 table, we calculate:

$$df = (r-1)(c-1) \quad \textbf{Formula 20.8}$$

which means: multiply the number of rows, minus one, by the number of columns, minus one. For table 20.6, there are:

$$(4-1 \text{ rows})(3-1 \text{ columns}) = 6 \text{ degrees of freedom}$$

Without even looking it up in appendix D (the χ^2 distribution), it's clear that the competing religions are not evenly distributed across the groups. If you

TABLE 20.6
A Hypothetical Census of Religious Belief in Four Groups of Native Americans

Tribe	Religion			Total
	Catholic	Protestants	Native American Church	
Observed frequencies				
1	150	104	86	340
2	175	268	316	759
3	197	118	206	521
4	68	214	109	391
Total	590	704	717	2,011
Expected frequencies				
1	99.75	119.03	121.22	
2	222.68	265.71	270.61	
3	152.85	182.39	185.76	
4	114.71	136.88	139.41	

had collected these data, you'd now be faced with the problem of interpreting them—that is, telling the story of how the various religions gain adherents at the expense of the others in various places.

Suppose that instead of a census, we take a 10% random sample of the groups—one that turns out to reflect almost perfectly the religious preferences of the population. The results, and the expected frequencies, would look like table 20.7.

Chi-square for table 20.7 is 15.44—still significant at the .02 level, but if we'd taken a 5% sample, χ^2 would be around 7 or 8, and with six degrees of freedom it would no longer be statistically significant, even at the .10 level. Sample size makes a real difference here.

The Special Case of the 2 × 2 Table

When you have a 2 × 2 cross-tab of nominal data, there is an easy formula to follow for computing χ^2 . Here it is:

$$\chi^2 = \frac{n(|ad - bc| - n/2)^2}{(a + b)(c + d)(a + c)(b + d)} \quad \text{Formula 20.9}$$

where a , b , c , and d are the individual cells shown in figure 20.2, and n is the total of all the cells (the lower-right-hand marginal). The four sums in the

TABLE 20.7
Observed and Expected Frequencies for a 10% Sample of the Data in Table 20.6

<i>Tribe</i>	<i>Religion</i>			<i>Total</i>
	<i>Catholic</i>	<i>Protestants</i>	<i>Native American Church</i>	
Observed frequencies				
1	15	10	9	34
2	18	27	32	77
3	20	12	21	53
4	7	21	11	39
Total	60	70	73	203
Expected frequencies				
1	10.04926	11.72414	121.22	
2	22.75862	26.55172	270.61	
3	15.66502	18.27586	185.76	
4	11.52709	11.52709	139.41	

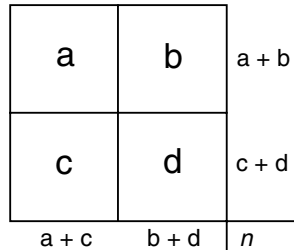


Figure 20.2. The cells in a 2×2 table.

denominator are the sums of the two rows and the two columns in figure 20.2. This is a good rough-and-ready way to calculate χ^2 quickly in the field if you don't have your data in a computer.

Marlene Dobkin de Rios (1981) studied the clientele of a Peruvian folk healer. She suspected that women clients were more likely to have had personal experience with witchcraft (or to have a close family member who has had personal contact with witchcraft) than were men clients. Table 20.8 shows Dobkin de Rios's data, and the calculation of chi-square directly from the cell frequencies, using formula 20.9. Dobkin de Rios's suspicion is certainly supported by these data.

TABLE 20.8
Dobkin de Rios’s Data on Experience with Witchcraft by Gender

	<i>Women clients</i>	<i>Men clients</i>	<i>Row totals</i>
Personal experience with witchcraft	63	12	75
No personal experience with witchcraft	5	15	20
Column totals	68	27	95

χ^2 using formula 20.9

$$\chi^2 = \frac{95(945 - 60)^2}{(75)(20)(68)(27)} = \frac{74,406,375}{2,754,000} = 27.02 \quad p < .001$$

SOURCE: Reprinted from *Social Science and Medicine*, Vol. 15B, M. Dobkin de Rios “Socioeconomic Characteristics of an Amazon Urban Healer’s Clientele,” pp. 51–63, © 1981, with permission of Elsevier.

The Odds Ratio

The **odds ratio**, while also not a PRE statistic, is a measure of the relation between nominal variables. In table 20.8, among those who report having personal experience with witchcraft the odds are 5.25-to-1 that it’s a woman ($63/12 = 5.25$). Among those who report not having had personal experience with witchcraft, the odds are one-third ($5/15 = .33$) that it’s a man.

For 2×2 tables, the two primary odds are a/c and b/d . The odds ratio is

$$\frac{a/c}{b/d} \qquad \textbf{Formula 20.10}$$

which can be collapsed to simply ad/bc . The odds ratio for is table 20.8 is:

$$945/60 = 15.75$$

If the ratio of a/c were the same as the ratio of b/d , then the odds ratio would be 1. The degree to which the odds ratio is not 1 is a measure of association between the variables. In this case, the ratio of a/c is more than 15 times bigger than the ratio of b/d . This means that women clients are that much more likely than men clients to have had personal experience with witchcraft.

Fisher’s Exact Test

Notice that there are only five cases in one of the cells in table 20.8—women who report no personal experience with witchcraft. When there are

fewer than five cases *expected* in any cell, χ^2 can be distorted and there is a danger of drawing the wrong conclusion from the data.

There are, however, $68(20)/95 = 14.32$ cases *expected* of women who report not having any personal experience with witchcraft. The fact that only five cases actually turned up in this cell is thus strong evidence, not a cause for worry.

When the *expected* number of frequencies for any cell in a 2×2 table is less than 5, use Fisher's exact probability test. Here's an example.

Table 20.9 shows data from Thomas Gregor's study (1979) of height and social success among the Mehinaku of the Brazilian Amazon. The median height of 15 unmarried, adult men in the tribe was 161.85 cm (just under 5'4"). The median number of girlfriends that each man had was four. The independent variable in table 20.9, then, is height, measured as equal to or less than

TABLE 20.9
Data on the Number of Girlfriends by Height

	≤ 161.85	> 161.85 cm	Column totals
≤ 4 girlfriends	7	2	9
[expected]	[4.8]	[4.2]	
> 4 girlfriends	1	5	6
[expected]	[3.2]	[2.8]	
Row totals	8	7	15

SOURCE: T. Gregor, "Short People." Reprinted from *Natural History*, Vol. 88, no. 2, pp. 14–21. Copyright © Natural History Magazine, Inc., 1979. Reproduced by permission.

the median, or greater than the median. The dependent variable is the number of girlfriends, measured as equal to or less than the median, or greater than the median.

Of the eight men who were at or below median height, seven had four or fewer girlfriends. Of the seven men who were above median height, five had more than four girlfriends. The *expected* values for each cell are shown in brackets in table 20.9.

Since the cells have expected values of less than 5, we apply Fisher's exact text. There are thousands of ways to throw the expected cases (5, 4, 3, 3) into four cells, but there are fewer ways to do it if you have to make the right-hand marginals add up to the observed values. Given a set of fixed marginals in a 2×2 table, the probability of seeing any distribution of cases is:

$$p = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{n!a!b!c!d!} \quad \text{Formula 20.11}$$

where a , b , c , and d are the actual contents of the four cells in the 2×2 table (see figure 20.2 above) and n is the total number of cases. An exclamation point signifies the factorial of a number, or the product of a sequence of numbers. So, $4!$ is four-factorial, or $4 \times 3 \times 2 \times 1 = 24$. Calculators can't handle the factorials of large numbers, but you can do the needed calculations on a spreadsheet when you have the relatively small numbers involved in a Fisher's exact test.

The exact probability of observing the particular distribution of cases in table 20.9 is:

$$p = \frac{(362,880)(720)(40,320)(5,040)}{(1,307,674,368,000)(5,040)(2)(1)(120)} = 0.033566434$$

For a one-tailed test of the null hypothesis, we need to add this probability to the probability of finding any other distributions that are more extreme than the one in the actual data. There is one configuration of these data that is more extreme: $a = 8$, $b = 0$, $c = 1$, and $d = 6$ (where a , b , c , and d are the cells in figure 20.2). The exact probability of that distribution is 0.001398601. For a one-tailed test, we add these probabilities:

$$0.033566434 + 0.001398601 = 0.034965$$

A two-tailed test is simply twice this probability, or 0.06993. There is a strong hint in these data that men who are above median height have more girlfriends than are men who are below median height in this tribe. But the case is tentative and awaits further tests on larger samples.

Gamma: The All-Purpose PRE Measure of Association for Ordinal Variables

Once you understand the PRE principle, a lot of things in statistics fall into place. Kempton et al. (1995) surveyed intentionally selected samples of people in the United States whom they thought would show pro- and anti-environmentalist attitudes. (Members of the Sierra Club, for example, are people you'd anticipate would be pro-environmental activism, while loggers are people you'd think would be against that kind of activity.) Here are two questions from Kempton et al.'s study:

You shouldn't force people to change their lifestyle for the sake of the environment.

1. Disagree
2. Neutral
3. Agree

Environmentalists wouldn't be so gung-ho if it were *their* jobs that were threatened.

1. Disagree 2. Neutral 3. Agree

Gery Ryan, Stephen Borgatti, and I used these items in a survey we ran (Bernard et al. 1999), and table 20.10 shows the results. Notice that these

TABLE 20.10
Distribution of Responses on Two Ecological Attitude Items

<i>Force change in lifestyle</i>	<i>Gung-ho</i>			<i>Row totals</i>
	<i>Disagree</i>	<i>Neutral</i>	<i>Agree</i>	
Disagree	16	7	61	84
Neutral	7	7	7	21
Agree	13	4	27	44
Column totals	36	18	95	149

items are reverse scored, so that a higher number indicates support for environmentalism.

If the two variables were perfectly related, then every respondent who agreed with one statement would agree with the other; and every respondent who disagreed with one statement would disagree with the other. Of course, things never work out so neatly, but if you knew the **proportion of matching pairs** among your respondents, you'd have a PRE measure of association for ordinal variables. The measure would tell you how much more correctly you could guess the rank of one ordinal variable for each respondent if you knew the score for the other ordinal variable in a bivariate distribution.

What we would like is a PRE measure of association that tells us whether knowing the ranking of pairs of people on one variable increases our ability to predict their ranking on a second variable, and by how much. To do this, we need to understand the ways in which pairs of ranks can be distributed. This will not appear obvious at first, but bear with me.

The number of possible pairs of observations (on any given unit of analysis) is

$$\text{No. of pairs of observations} = \frac{n(n-1)}{2} \qquad \textbf{Formula 20.12}$$

where n is the sample size. There are $(149)(148)/2 = 11,026$ pairs of observations in table 20.10.

There are several ways that pairs of observations can be distributed if they are ranked on two ordinal variables.

1. They can be ranked in the same order on *both* variables. We'll call these "same."
2. They can be ranked in the opposite order on both variables. We'll call these "opposite."
3. They can be tied on either the independent or dependent variables, or on both. We'll call these "ties."

In fact, in almost all bivariate tables comparing ordinal variables, there are going to be a lot of pairs with tied values on both variables. **Gamma**, written G , is a popular measure of association between two ordinal variables because it *ignores* all the tied pairs. The formula for gamma is:

$$G = \frac{\text{No. of same-ranked pairs} - \text{No. of opposite-ranked pairs}}{\text{No. of same-ranked pairs} + \text{No. of opposite-ranked pairs}} \quad \text{Formula 20.13}$$

Gamma is an intuitive statistic; it ranges from -1.0 (for a perfect negative association) to $+1.0$ (for a perfect positive association), through 0 in the middle for complete independence of two variables.

If there are just two ordinal ranks in a measure, and if the number of opposite-ranked pairs is 0 , then gamma would equal 1 . Suppose we measure income and education in a Malaysian community ordinaly, such that: (1) Anyone with less than a high school diploma is counted as having low education, and anyone with at least a high school diploma is counted as having high education; and (2) Anyone with an income of less than RM34,000 a year is counted as having low income, while anyone with at least RM34,000 a year is counted as having high income. (RM is the symbol for the Malaysian ringgit. RM34,000, or about \$9,000 U.S. dollars, was the average per capita GDP in Malaysia in 2005.)

Now suppose that *no one* who had at least a high school diploma earned less than RM34,000 a year. There would be no pair of observations, then, in which low income and high education (an opposite pair) co-occurred.

If the number of same-ranked pairs is zero, then gamma would equal $+1.0$. Suppose that *no one* who had high education also had a high income. This would be a perfect negative association, and gamma would be -1.0 . Both $+1.0$ and -1.0 are perfect correlations.

Calculating the Pairs for Gamma

The number of same-ranked pairs in a bivariate table is calculated by multiplying each cell by the sum of all cells *below it and to its right*. The number

of opposite-ranked pairs is calculated by multiplying each cell by the sum of all cells *below it and to its left*. This is diagramed in figure 20.3.

<p>In Table 20.10, the number of same-ranked pairs is:</p> $ \begin{array}{r} 16 (7 + 7 + 4 + 27) = 720 \\ + 7 (7 + 27) = 238 \\ + 7 (4 + 27) = 217 \\ + 7 (27) = 189 \\ \hline \text{Total} \qquad \qquad 1,364 \end{array} $	<p>The number of opposite-ranked pairs is:</p> $ \begin{array}{r} 61 (7 + 7 + 13 + 4) = 1,891 \\ + 7 (7 + 13) = 140 \\ + 7 (4 + 13) = 119 \\ + 7 (13) = 91 \\ \hline \text{Total} \qquad \qquad 2,241 \end{array} $
---	--

Figure 20.3. Calculating gamma.

Gamma for table 20.10, then, is:

$$G = \frac{1,364 - 2,241}{1,364 + 2,241} = \frac{-877}{3,605} = -.24$$

Gamma tells us that the variables are associated negatively—people who agree with either of the statements tend to disagree with the other, and vice versa—but it also tells us that the association is relatively weak.

Is Gamma Significant?

How weak? If you have more than 50 elements in your sample, you can test for the probability that gamma is due to sampling error using a procedure developed by Goodman and Kruskal (1963). A useful presentation of the procedure is given by Loether and McTavish (1993:598, 609). First, the gamma value must be converted to a z-score, or standard score. The formula for converting gamma to a z-score is:

$$z = (G - \gamma) \sqrt{(n_s + n_o) / 2n(1 - G^2)} \quad \textbf{Formula 20.14}$$

where G is the *sample* gamma, γ is the gamma for the *population*, n is the size of your sample, n_s is the number of same-ranked pairs, and n_o is the number of opposite-ranked pairs.

As usual, we proceed from the null hypothesis and assume that γ for the entire population is zero—that is, that there really is no association between the variables we are studying. If we can reject that hypothesis, then we can assume that the gamma value for our sample probably approximates the gamma value, γ , for the population. Using the data from figure 20.3 and the gamma value for table 20.10:

$$z = (-.24 - 0) \sqrt{\frac{1,364 + 2,241}{2(149)(1 - (-.24^2))}}$$

You'll recall that appendix B—the *z*-score table—lists the proportions of area under a normal curve that are described by various *z*-score values. To test the significance of gamma, look for the *z*-score in column 1 of the table. Column 2 shows the area under a normal curve between the mean (assumed to be zero for a normal curve) and the *z*-score. We're interested in column 3, which shows the area under the curve that is *not* accounted for by the *z*-score.

A *z*-score of $-.86$ accounts for all but $.1949$ of the area under a normal curve. This means that we can *not* reject the null hypothesis. The gamma score of $-.24$ is not sufficiently strong to confirm that there is a significant association between responses to the two attitudinal questions about environmental activism.

Kendall's T_b

Some researchers prefer a statistic called **Kendall's T_b** (pronounced tau-b and also written τ_b) instead of gamma for bivariate tables of ordinal data because gamma ignores tied pairs in the data. The formula for T_b is:

$$T_b = \frac{n_s - n_o}{\sqrt{(n_s + n_o + n_{td})(n_s + n_o + n_{ti})}} \quad \textbf{Formula 20.15}$$

where n_s is the number of same-ranked pairs, n_o is the number of opposite-ranked pairs, n_{td} is the number of pairs tied on the dependent variable, and n_{ti} is the number of pairs tied on the independent variable. The formula for calculating the tied pairs is:

$$n_{td} = \sum R(R - 1)/2 \quad n_{ti} = \sum C(C - 1)/2 \quad \textbf{Formula 20.16}$$

where R refers to the row marginals (the dependent variable) and C refers to the column marginals (the independent variable). In table 20.10:

$$n_{td} = [84(83) + 21(20) + 44(43)]/2 = 4,642$$

$$n_{ti} = [36(35) + 18(17) + 95(94)]/2 = 5,248$$

We already have the numerator for T_b in this case (we calculated the number of same-ranked and opposite-ranked pairs in figure 20.3), so:

$$T_b = \frac{1,364 - 2,241}{\sqrt{(1,364 + 2,241 + 4,462)(1,364 + 2,241 + 5,248)}} = -.10$$

This confirms the weak, negative association we saw from the results of the gamma test. Kendall's T_b will usually be smaller than gamma because gamma ignores tied pairs, while T_b uses almost all the data (it ignores the relatively few pairs that are tied on both variables).

Yule's Q : Gamma for 2×2 Tables

Yule's Q is the equivalent of gamma for 2×2 tables of ordinal variables, like high vs. low prestige, salary, education, religiosity, and so on. Yule's Q can be calculated on frequencies or on percentages. The formula is:

$$Q = \frac{ad - bc}{ad + bc} \quad \text{Formula 20.17}$$

Yule's Q is a handy, easy-to-use statistic. A good rule of thumb for interpreting Q is given by Davis (1971): When Q is 0, the interpretation is naturally that there is no association between the variables. When Q ranges between from 0 to $-.29$, or from 0 to $+.29$, you can interpret this as a negligible or small association. Davis interprets a Q value of $\pm .30$ to $\pm .49$ as a "moderate" association; a value of $\pm .50$ to $\pm .69$ as a "substantial" association; and a value of $\pm .70$ or more as a "very strong" association.

Rutledge (1990) was interested in the effect of one- or two-parent families on children's relations with their mothers and fathers. She surveyed African American, college-aged women, mostly from Chicago. One of the questions she asked was: "When you were growing up, how close were you to your father? Were you considerably close, moderately close, or not close at all?" I've collapsed Rutledge's data into two response categories, close and not close, in table 20.11.

TABLE 20.11
Family Structure and Self-Reported Closeness to Parents

<i>Close to father?</i>	<i>Two Parents</i>	<i>One Parent</i>	<i>Total</i>
Yes	135	36	171
No	13	31	44
Total	148	67	215

SOURCE: E. M. Rutledge, "Black Parent-Child Relations: Some Correlates," *Journal of Comparative Family Studies*, Vol. 21, pp. 369-378, 1990. Reprinted by permission.

Here is the calculation of Yule's Q for these data:

$$Q = \frac{(135)(31) - (36)(13)}{(135)(31) + (36)(13)} = \frac{4,185 - 468}{4,185 + 468} = .80$$

Yule's Q for these data is .80. Most of the women ($135/148 = .91$) who come from two-parent homes are close to their fathers, compared to fewer than half who come from one-parent homes ($31/67 = .46$). The reason is obvious: Overwhelmingly, one-parent homes are headed by mothers, not by fathers.

What to Use for Nominal and Ordinal Variables

In general:

1. Use χ^2 to see how often you could expect to find the differences you see in the table just by chance. Calculate odds ratios to measure the strength of relationships.
2. Use gamma (or tau, or—in the case of 2×2 tables—Yule's Q) to measure the association between two ordinal variables.

In actual practice, ordinal variables with seven ranks are treated as if they were interval variables. In fact, many researchers treat ordinals with just five ranks as if they were intervals, because association between interval-level variables can be analyzed by the most powerful statistics—which brings us to correlation and regression.

Correlation: The Powerhouse Statistic for Covariation

When at least one of the variables in a bivariate relation is interval or ratio level, we use a measure of correlation: **Spearman's r** , written r_s when the data are rank ordered; **Pearson's product moment correlation**, written simply as r , to measure the strength of linear relations; or **eta squared** (eta is the Greek letter η , pronounced either eat-a or ate-a) to measure the strength of certain kinds of nonlinear relations. (Go back to the section on "shape of relations" at the beginning of this chapter if you have any doubts about the concept of a nonlinear relation.)

Spearman's r

Table 20.12 shows some data collected by Allyn Stearman (1989:224). She measured the amount of game meat taken, over 56 days, by 16 Yuquí hunters in the Bolivian Amazon.

These data are arranged in rank order of meat taken. So, Alejandro is the winner, with 226 kilos of meat during the 56-day observation period. Lorenzo

TABLE 20.12
Ranking of Yuquí Hunters by Game Taken

<i>Name</i>	<i>Kilos of meat</i>	<i>Kilos of fish</i>
Alejandro	226.00	53.75
Jaime	185.50	101.00
Leonardo	152.50	8.50
Humberto	144.75	120.50
Daniel	78.00	119.50
Joel	74.50	34.25
Jorge	59.50	23.00
Timoteo	51.00	1.50
Tomás	51.00	123.80
Lucas	46.00	107.50
Guillermo	45.75	199.25
Victor	29.50	38.25
Manuel	14.50	28.50
Benjamin	10.00	128.00
Jonatán	0.00	198.00
Lorenzo	0.00	279.00

SOURCE: A. M. Stearman, "Yuquí Foragers in the Bolivian Amazon: Subsistence Strategies, Prestige and Leadership in an Acculturating Society," *Journal of Anthropological Research*, Vol. 45, pp. 219–244. Reproduced by permission of the *Journal of Anthropological Research* © 1989.

and Jonatán are tied for last. We can use Spearman's r to test the extent to which a hunter's take of fish matches his take of meat. Just looking at the data, we suspect the correlation is negative. That is, it appears that hunters who take a lot of meat are not focusing on fish and vice versa. Here is the formula for Spearman's r :

$$r_s = 1 - \frac{6\sum d^2}{n(n^2 - 1)} \quad \text{Formula 20.18}$$

where d is the difference between the ranks on all pairs of objects. Table 20.13 shows the computation of d^2 for the data on these hunters.

For these hunters, at least, there is a moderate negative correlation between their success in hunting meat and their success in hunting fish.

Pearson's r

Pearson's r measures how much changes in one variable correspond with equivalent changes in the other variables. It can also be used as a measure of

TABLE 20.13

Computing Spearman’s Rank Order Correlation Coefficient for the Data in Table 20.12

<i>Hunter</i>	<i>Rank for meat</i>	<i>Rank for fish</i>	<i>Difference in the ranks</i>	<i>d²</i>
Alejandro	1	10	−9	81
Jaime	2	9	−7	49
Leonardo	3	15	−12	144
Humberto	4	6	−2	4
Daniel	5	7	−2	4
Joel	6	12	−6	36
Jorge	7	14	−7	49
Timoteo	8	16	−8	64
Tomás	9	5	4	16
Lucas	10	8	2	4
Guillermo	11	2	9	81
Victor	12	11	1	1
Manuel	13	13	0	0
Benjamín	14	4	10	100
Jonatán	15	3	12	144
Lorenzo	16	1	15	225
				total <i>d</i> ² 1,002

$$r_s = 1 - \frac{6(1002)}{16(16^2 - 1)} = 1 - 6012/4080 = -.474$$

association between an interval and an ordinal variable, or between an interval and a **dummy variable**. (Dummy variables are nominal variables coded as 1 or 0, present or absent. See chapter 17 on text analysis.) The square of Pearson’s *r*, or **r-squared**, is a PRE (proportionate reduction of error) measure of association for linear relations between interval variables. It tells us how much better we could predict the scores of a dependent variable, if we knew the scores of some independent variable.

Table 20.14 shows data for two interval variables for a random sample of 10 of the 50 countries in table 19.8: (1) infant mortality and (2) life expectancy for women.

To give you an idea of where we’re going with this example, the correlation between INFMORT and TFR across the 50 countries in table 19.8 is around .89, and this is reflected in the sample of 10 countries for which the correlation is *r* = .94.

Now, suppose you had to predict the TFR for each of the 10 countries in table 20.14 *without knowing anything about the infant mortality rate for those*

TABLE 20.14
Infant Mortality by TFR for 10 Countries from Table 19.8

<i>State</i>	<i>INFMORT</i> <i>x</i>	<i>TFR</i> <i>y</i>
Armenia	26	1.70
Chad	112	6.07
El Salvador	32	3.17
Ghana	66	5.15
Iran	35	2.80
Latvia	18	1.25
Namibia	65	4.90
Panama	21	2.63
Slovenia	7	1.26
Suriname	29	2.21
Mean of $x = 41.10$		Mean of $y = 3.114$

countries. Your best guess—your lowest prediction error—would be the mean, 3.114 children per woman. You can see this in figure 20.4 where I’ve plotted the distribution of TFR and INFMORT for the 10 countries in table 20.14.

The Sums of the Squared Distances to the Mean

Each dot in figure 20.4 is physically distant from the dotted mean line by a certain amount. The sum of the squares of these distances to the mean line is the smallest sum possible (that is, the smallest cumulative prediction error you could make), given that you *only* know the mean of the dependent variable. The distances from the dots *above* the line to the mean are positive; the distances from the dots *below* the line to the mean are negative. The sum of the actual distances is zero. Squaring the distances gets rid of the negative numbers.

But suppose you *do* know the data in table 20.14 regarding the infant mortality rate for each of those 10 countries. Can you reduce the prediction error in guessing the TFR for those countries? Could you draw another line through figure 20.4 that “fits” the dots better and reduces the sum of the distances from the dots to the line?

You bet you can. The solid line that runs diagonally through the graph in figure 20.4 minimizes the prediction error for these data. This line is called the **best fitting line**, or the **least squares line**, or the **regression line**. When you understand how this regression line is derived, you’ll understand how correlation works.

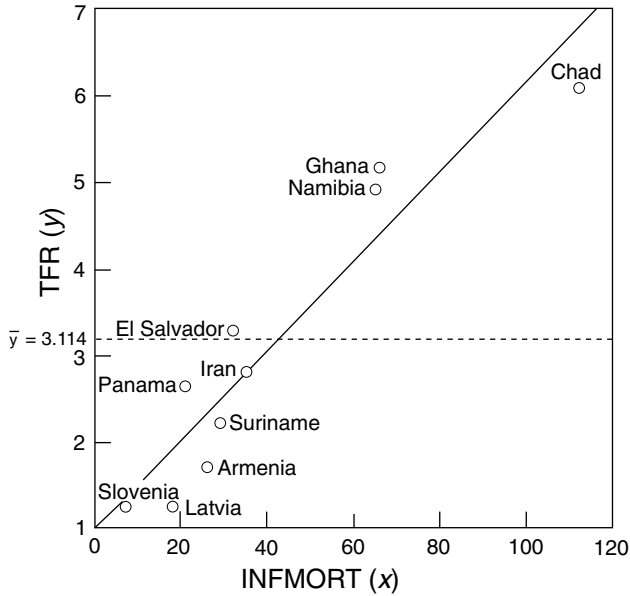


Figure 20.4. A plot of the data in table 20.14. The dotted line is the mean of TFR. The solid line is drawn from the regression equation $y = 1.018 + .051x$.

Regression

The formula for the regression line is

$$y = a + bx \quad \text{Formula 20.19}$$

where y is the variable value of the dependent variable, a and b are some constants (which you'll learn how to derive in a moment), and x is the variable value of the independent variable. The constant, a , is computed as:

$$a = \bar{y} - b\bar{x} \quad \text{Formula 20.20}$$

and b is computed as

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} \quad \text{Formula 20.21}$$

Table 20.15 shows the data needed for finding the regression equation for the raw data in table 20.14. At the bottom of table 20.15 you'll find a shortcut formula (Formula 20.22) for calculating Pearson's r directly from the data in the table.

TABLE 20.15
Computation of Pearson's r Directly from Data in Table 20.14

<i>Country</i>	<i>INFMORT</i> x	<i>TFR</i> y	xy	x^2	y^2
Armenia	26	1.70	44.20	676	2.890
Chad	112	6.07	679.84	12,544	36.845
El Salvador	32	3.17	101.44	1,024	10.049
Ghana	66	5.15	339.90	4,356	26.523
Iran	35	2.80	98.00	1,225	7.840
Latvia	18	1.25	22.50	324	1.562
Namibia	65	4.90	318.50	4,225	24.010
Panama	21	2.63	55.23	44	6.917
Slovenia	7	1.26	8.82	49	1.588
Suriname	29	2.21	64.09	841	4.884
total	411	31.14	1,732.52	25,705	123.107
$N = 10$	$\bar{x} = 41.1$	$\bar{y} = 3.114$			

$$r_{xy} = \frac{n\sum xy - \sum x \sum y}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

Formula 20.22

$$r_{xy} = \frac{10(1,732.52) - (411)(31.14)}{\sqrt{(257,050 - 168,924)(1,231.07 - 969.7)}} = .943$$

The constant b is:

$$b = \frac{10(1,732.52) - (411)(31.14)}{10(25,705) - (411)^2} = \frac{4,526.66}{88,129} = 0.051$$

and the constant a is then:

$$a = 3,114 - .051(41.1) = 1.018$$

The regression equation for any pair of scores on infant mortality (x) and TFR (y), then, is:

$$y = a + bx = 1.018 + .05(x)$$

Drawing the Regression Line

To draw the regression line in figure 20.4, find the expected y coordinate for two of the actual data points on the independent variable, x . Choose the x val-

ues that are farthest apart to maximize the accuracy of the line. For table 20.16, choose Slovenia and Chad:

1. For Slovenia, the expected TFR is:

$$y = 1.018 + .051(7) = 1.375$$

2. For Chad, the expected TFR is:

$$y = 1.018 + .051(112) = 6.730$$

Put a dot on figure 20.4 at the intersection of

$$x = 7, \quad y = 1.375$$

and

$$x = 112, \quad y = 6.730$$

and connect the dots. That's the regression line.

The squared deviations (the distances from any dot to the line, squared) add up to less than they would for any other line we could draw through that graph. That's why the regression line is also called the **best fitting** or the **least squares** line.

Suppose we want to predict the dependent variable y (TFR) when the independent variable x (INFMORT) is 4, as it is in the Czech Republic. In that case,

$$y = 1.018 + .051(4) = 1.22$$

or 1.22 total births during the lifetime of women born between 1995 and 2000 in Czech Republic (the current TFR for the Czech Republic is, in fact, about 1.2). In other words, the regression equation lets us estimate the TFR for infant mortality levels that are not represented in our sample.

How Regression Works

To give you an absolutely clear idea of how the regression formula works, table 20.16 shows all the predictions along the regression line for the data in table 20.14.

We now have two predictors of TFR: (1) the mean TFR, which is our best guess when we have no data about some independent variable like infant mortality and (2) the values produced by the regression equation when we *do* have information about something like infant mortality.

Each of these predictors produces a certain amount of error, or **variance**,

TABLE 20.16
Regression Predictions for the Dependent Variable in Table 20.14

<i>For the country of</i>	<i>where the infant mortality rate in 2000 was</i>	<i>predict that the TFR will be</i>	<i>and compare that to the actual TFR in table 19.8</i>
Armenia	26	$1.018 + .051(26) = 2.344$	1.70
Chad	112	$1.018 + .051(112) = 6.730$	6.07
El Salvador	32	$1.018 + .051(32) = 2.650$	3.17
Ghana	66	$1.018 + .051(66) = 4.384$	5.15
Iran	35	$1.018 + .051(35) = 2.803$	2.80
Latvia	18	$1.018 + .051(18) = 1.936$	1.25
Namibia	65	$1.018 + .051(65) = 4.333$	4.90
Panama	21	$1.018 + .051(21) = 2.089$	2.63
Slovenia	7	$1.018 + .051(7) = 1.375$	1.26
Suriname	29	$1.018 + .051(29) = 2.497$	2.21

which is the difference between the predicted number for the dependent variable and the actual measurement. This is also called the **residual**—that is, what’s left over after making your prediction using the regression equation. (To anticipate the discussion of multiple regression in chapter 21: The idea in multiple regression is to use two or more independent variables in order to reduce the size of the residuals.)

You’ll recall from chapter 19, in the section on variance and the standard deviation, that in the case of the mean, the total variance is the average of the squared deviations of the observations from the mean, $\left[\sum (x - \bar{x})^2 / n \right]$. In the case of the regression line predictors, the variance is the sum of the squared deviations from the regression line. Table 20.17 compares these two sets of errors, or variances, for the data in table 20.14.

We now have all the information we need for a true PRE measure of association between two interval variables. Recall the formula for a PRE measure: the old error minus the new error, divided by the old error. For our example in table 20.14:

$$\text{PRE} = \frac{26.136 - 2.890}{26.136} = .889$$

In other words: The proportionate reduction of error in guessing the TFR in table 20.14—given that you know the distribution of informant mortality rates and can apply a regression equation—compared to just guessing the mean of TFR is .889, or about 89%.

TABLE 20.17
 Comparison of the Error Produced by Guessing the Mean TFR in Table 20.14 and the Error Produced by Applying the Regression Equation for Each Guess

Country	TFR <i>y</i>	Old Error $(y - \bar{y})^2$	Prediction using the regression equivalent	New Error $(y - \text{the prediction using the regression equation})^2$
Armenia	1.70	1.999	2.344	.415
Chad	6.07	8.738	6.730	.436
El Salvador	3.17	.003	2.650	.270
Ghana	5.15	4.145	4.384	.587
Iran	2.80	.099	2.803	.000
Latvia	1.25	3.474	1.936	.470
Namibia	4.90	3.190	4.333	.321
Panama	2.63	.234	2.089	.293
Slovenia	1.26	3.437	1.375	.013
Suriname	2.21	.817	2.497	.082
		$\Sigma = 26.136$		$\Sigma = 2.890$

This quantity is usually referred to as **r-squared** (written r^2), or the amount of variance accounted for by the independent variable. It is also called the **coefficient of determination** because it tells us how much of the variance in the dependent variable is predictable from (determined by) the scores of the independent variable. The Pearson product moment correlation, written as r , is the square root of this measure, or, in this instance, .943. We calculated r in table 20.15 by applying formula 20.22.

Calculating r and r^2

I've given you this grand tour of regression and correlation because I want you to see that Pearson's r is not a direct PRE measure of association; its square, r^2 , is.

So, what's better, Pearson's r or r^2 for describing the relation between interval variables? Pearson's r is easy to compute from raw data and it varies from -1 to $+1$, so it has direction and an intuitive interpretation of magnitude. Mathematically, of course, r has to be at least as big (and almost always bigger) than r^2 . By contrast, r^2 is a humbling statistic. A correlation of .30 looks impressive until you square it and see that it explains just 9% of the variance in what you're studying.

The good news is that if you double a correlation coefficient, you quadruple the variance accounted for. For example, if you get an r of .25, you've accounted for 6.25% of the variance, or error, in predicting the score of a dependent variable from a corresponding score on an independent variable. An r of .50 is twice as big as an r of .25, but four times as good, because $.50^2$ means that you've accounted for 25% of the variance.

Accounting for Variance

What does “accounting for variance” mean? It simply means being able to make a better prediction with an independent variable than we could without one. Recall from chapter 19 that the total amount of variance in an observed set of scores is the sum of the squared deviations from the mean, divided by $n - 1$. This is shown in Formula 20.23.

$$\text{Variance} = \frac{\sum (x - \bar{x})^2}{n - 1} \quad \text{Formula 20.23}$$

Well, if the correlation between the two variables in table 20.14 is $r = 0.943$, and its square is $r^2 = 0.889$, then we can say “we have accounted for 89% of the variance.” I'll define what I mean by this.

The mean TFR for the 10 countries in table 20.15 is 3.114, and the total variance (sum of squares) is 26.136. When we predict TFR, using the regression formula, from the infant mortality rate, the variance in the predicted scores will be

$$(1 - .889)26.136 = 2.90$$

Except for rounding error, this is the total “new error” in table 20.17 and this is what it means to say that “about 89% of the variance in TFR is [accounted for] [determined by] [predicted by] the infant mortality rate across 10 countries.”

By the way, in case you're wondering, the correlation between infant mortality and TFR across the 50 countries in table 19.8 is .89 and the correlation between these two variables across all countries of the world is .86.

Testing the Significance of r

Just as with gamma, it is possible to test whether or not any value of Pearson's r is the result of sampling error, or reflects a real covariation in the larger

population. In the case of r , the null hypothesis is that, within certain confidence limits, we should predict that the real coefficient of correlation in the population of interest is actually zero. In other words, there is no relation between the two variables.

We need to be particularly sensitive to the possible lack of significance of sample statistics when we deal with small samples—which is a lot of the time, it turns out. To simplify testing the confidence limits of r , I have constructed table 20.18, which you can use to get a ballpark reading on the significance of Pearson's r . The top half of table 20.18 shows the 95% confidence limits for representative samples of 30, 50, 100, 400, and 1,000, where the Pearson's r

TABLE 20.18
Confidence Limits for Pearson's r for Various Sample Sizes

Pearson's r	Sample Size				
	30	50	100	400	1,000
0.1	ns	ns	ns	ns	.04-.16
0.2	ns	ns	.004-.40	.10-.29	.14-.26
0.3	ns	.02-.54	.11-.47	.21-.39	.24-.35
0.4	.05-.67	.14-.61	.21-.55	.32-.48	.35-.45
0.5	.17-.73	.25-.68	.31-.63	.42-.57	.45-.54
0.6	.31-.79	.39-.75	.45-.71	.53-.66	.56-.64
0.7	.45-.85	.52-.82	.59-.79	.65-.75	.67-.73
0.8	.62-.90	.67-.88	.72-.86	.76-.83	.78-.82
0.9	.80-.95	.83-.94	.85-.93	.88-.92	.89-.91
Top half of table: 95% confidence limits					
0.1	ns	ns	ns	ns	.02-.18
0.2	ns	ns	ns	.07-.32	.12-.27
0.3	ns	ns	.05-.51	.18-.41	.23-.45
0.4	ns	.05-.80	.16-.59	.28-.50	.33-.46
0.5	.05-.75	.17-.72	.28-.67	.40-.59	.44-.56
0.6	.20-.83	.31-.79	.41-.74	.51-.68	.55-.65
0.7	.35-.88	.46-.85	.55-.81	.63-.76	.66-.74
0.8	.54-.92	.62-.90	.69-.88	.75-.84	.77-.83
0.9	.75-.96	.80-.95	.84-.94	.87-.92	.88-.91
Bottom half of table: 99% confidence limits					

values are .1, .2, .3, etc. The bottom half of table 20.18 shows the 99% confidence limits.

Reading the top half of table 20.18, we see that at the 95% level the confidence limits for a correlation of .20 in a random sample of 1,000 are .14 and

.26. This means that in fewer than five tests in a hundred would we expect to find the correlation smaller than .14 or larger than .26. In other words, we are 95% confident that the true r for the population (written ρ , which is the Greek letter rho) is somewhere between .14 and .26.

By contrast, the 95% confidence limits for an r of .30 in a random sample of 30 is not significant at all; the true correlation could be 0, and our sample statistic of .30 could be the result of sampling error.

The 95% confidence limits for an r of .40 in a random sample of 30 is statistically significant. We can be 95% certain that the true correlation in the population (ρ) is no less than .05 and no larger than .67. This is a statistically significant finding, but not much to go on insofar as external validity is concerned. You'll notice that with large samples (like 1,000), even very small correlations are significant at the .01 level.

On the other hand, just because a statistical value is significant doesn't mean that it's important or useful in understanding how the world works. Looking at the lower half of table 20.18, we see that even an r value of .40 is statistically insignificant when the sample is as small as 30. If you look at the spread in the confidence limits for both halves of table 20.18, you will notice something very interesting: A sample of 1,000 offers *some* advantage over a sample of 400 for bivariate tests, but the difference is small and the costs of the larger sample could be very high, especially if you're collecting all your own data.

Recall from chapter 7, on sampling, that in order to halve the confidence interval you have to quadruple the sample size. Where the unit cost of data is high—as in research based on direct observation of behavior or on face-to-face interviews—the point of diminishing returns on sample size is reached quickly. Where the unit cost of data is low—as it is with mailed questionnaires or with telephone surveys—a larger sample is worth trying for.

Nonlinear Relations

And now for something different. All the examples I've used so far have been for linear relations where the best-fitting "curve" on a bivariate scatterplot is a straight line. A lot of really interesting relations, however, are nonlinear.

Smits et al. (1998) measured the strength of association between the educational level of spouses in 65 countries and how that relates to industrialization. Figure 20.5 shows what they found. It's the relation between per capita energy consumption (a measure of industrialization and hence of economic development) and the amount of educational homogeneity in those 65 countries.

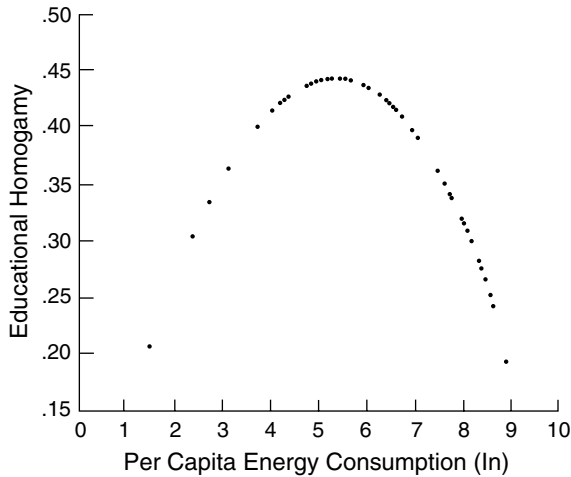


Figure 20.5. The relation between per capita energy consumption and educational homogamy is nonlinear.

SOURCE: J. Smits et al., “Educational Homogamy in 65 Countries: An Explanation of Differences in Openness Using Country-Level Explanatory Variables,” *American Sociological Review* Vol. 63, pp. 264–285. ©1998 by the American Sociological Association. Reprinted with permission.

(*Homogamy* means marrying someone who is similar to you, so educational homogamy means marrying someone who has the same level of education as you do.) The relation between the two variables is very clear: It’s an inverted U.

You might think that the relation between these two variables would be linear—people of similar education would be attracted to each other—but it isn’t. Here’s how Smits et al. reckon the inverted U happens: People in nonindustrialized countries rely mostly on agriculture, so family background (which determines wealth) tends to be the main criterion for selecting mates. As industrialization takes off, education becomes more and more important and family background becomes less important for building wealth. Educated people seek each other out to maximize their life chances.

Eventually, though, when countries are highly industrialized, education is widespread, people have high wages, and there are social security systems. People in *those* countries don’t have to rely on their children for support in old age, so romantic love becomes the dominant force in mate selection and the level of educational homogamy drops like a stone.

Here’s another one. Beginning with -5°F and continuing up to 75°F , the mean number of assaults rises steadily with the average daily temperature in

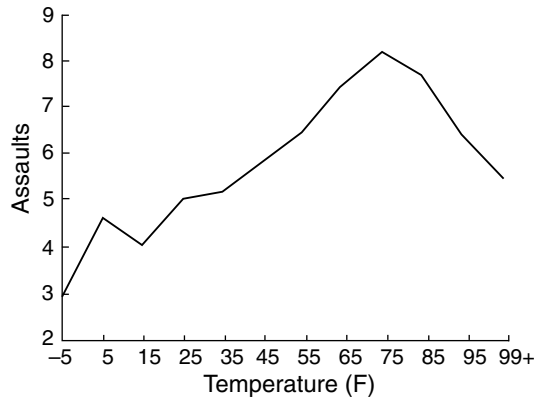


Figure 20.6. The relationship between temperature and the rate of assault is nonlinear. SOURCE: E. G. Cohn and J. Rotton, "Assault as a Function of Time and Temperature: A Moderator-Variable Time-Series Analysis," *Journal of Personality and Social Psychology*, Vol. 72, pp. 1322–1334. © 1997 by the American Psychological Association. Reprinted with permission.

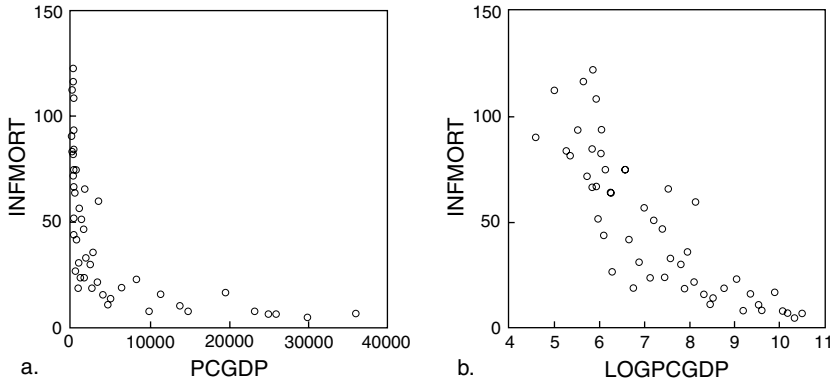
U.S. cities. Then it begins to drop. Figure 20.6, from Cohn and Rotton (1997), shows the dramatic pattern. The reasons for this pattern are very complex, but it is clear from figure 20.6 that a simple, linear correlation is inadequate to describe what's going on.

If you get a very weak r for two variables that you believe, from theory or from field research, are strongly related, then draw a scatterplot and check it out. Scatterplots are available in all the major statistical packages and they are, as you saw in figure 20.1, packed with information. For sheer intuitive power, there is nothing like them.

Figure 20.7a, for example, is the same plot we saw in figure 20.1d of infant mortality and per capita gross domestic product in 2000 for the 50 countries in table 19.8. I've repeated the figure here because I want you to see it next to figure 20.7b, which is a plot of infant mortality and the **natural logarithm** of PCGDP.

The correlation between the original variables in figure 20.7a is -0.60 . This indicates a negative relation, but the scatterplot clearly shows that the relation is not linear. Transforming each value of PCGDP into its natural logarithm has the effect of making the distribution more normal, but it changes the interpretation of the graph.

In figure 20.7b, the correlation is now a very strong -0.84 , but countries at the bottom, with \$100 dollars in per capita domestic product, have to raise their PCGDP by a factor of 10 in order to cut their infant mortality rate in half. At \$1,000 per year, countries only have to raise their per capita income



Infant mortality by per capita gross domestic product for 50 countries in table 19.8.

Infant mortality and the log of per capita gross domestic product for 50 countries in table 19.8.

Figure 20.7. Transforming variables.

by a factor of five to cut their infant mortality rate in half. At \$5,000 per year, countries only have to double PCGDP to halve infant mortality rate. And at \$10,000 per year, it only takes a 50% rise in PCGDP to halve the infant mortality rate again.

If a scatterplot looks anything like figure 20.7a, consider transforming the data into their logarithms. If a scatterplot looks like the shapes in figures 20.5 or 20.6, then consider using eta-squared, a statistic for nonlinear regression.

Calculating Eta-Squared

Eta-squared, written η^2 or eta^2 , is a PRE measure that tells you how much better you could do if you predicted the separate means for *chunks* of your data than if you predicted the mean for all your data. Figure 20.8 graphs the hypothetical data in table 20.19. These data show, for a sample of 20 people, ages 12–89, their “number of close friends and acquaintances.”

The dots in figure 20.8 are the data points from table 20.19. Respondent #10, for example, is 45 years of age and was found to have approximately 55 friends and acquaintances. The horizontal dashed line in figure 20.8 marked \bar{y} is the global average for these data, 41.15.

Clearly: (1) The global average is not of much use in predicting the dependent variable; (2) Knowing a person’s age *is* helpful in predicting the size of

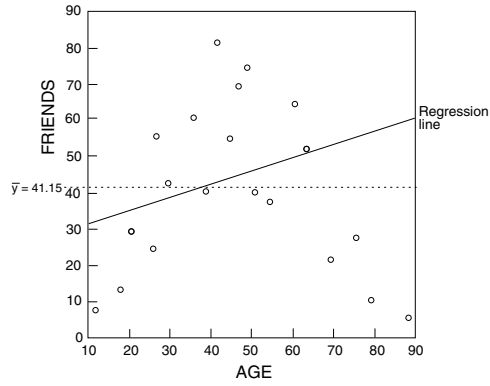


Figure 20.8. Number of friends by age.

TABLE 20.19
Hypothetical Data on Number of Friends by Age

<i>Person</i>	<i>Age</i>	<i>Number of Friends</i>	
1	12	8	
2	18	14	$\bar{y}_1 = 19.25$
3	21	30	
4	26	25	
5	27	56	
6	30	43	
7	36	61	
8	39	41	
9	42	82	$\bar{y}_2 = 57.00$
10	45	55	
11	47	70	
12	49	75	
13	51	41	
14	55	38	
15	61	65	
16	64	52	
17	70	22	
18	76	28	$\bar{y}_3 = 23.80$
19	80	11	
20	89	6	
			$\bar{y} = 41.15$

his or her social network; but (3) The linear regression equation is hardly any better than the global mean at reducing error in predicting the dependent variable. You can see this by comparing the mean line and the regression line (the slightly diagonal line in figure 20.8). They are very similar.

What that regression line depicts, of course, is the correlation between age and size of network, which is a puny $-.099$. But if we inspect the data visually, we find that there are a couple of natural “breaks.” It looks like there’s a break in the late 20s, and another somewhere in the 60s. We’ll break these data into three age chunks from 12 to 26, 27 to 61, and 64 to 89, take separate means for each chunk, and see what happens. I have marked the three chunks and their separate means on table 20.19.

Like r , eta must be squared to find the variance accounted for. η^2 is calculated from the following formula:

$$\eta^2 = 1 - \frac{\sum (y - \bar{y}_c)^2}{\sum (y - \bar{y})^2} \quad \text{Formula 20.24}$$

where \bar{y}_c is the average for each chunk and \bar{y} is the overall average for your dependent variable. For table 20.19, eta-squared is:

$$\eta^2 = 1 - \frac{3,871.55}{10,058.55} = .62$$

which is the proportionate reduction of error in predicting the number of friends people have from the three separate averages of their age, rather than from the global average of their age. This shows a pretty strong relation between the two variables, despite the very weak Pearson’s r .

Statistical Significance, the Shotgun Approach, and Other Issues

To finish this chapter, I want to deal with four thorny issues in social science data analysis: (1) measurement and statistical assumptions, (2) eliminating the outliers, (3) significance tests, and (4) the shotgun method of analysis.

Measurement and Statistical Assumptions

By now you are comfortable with the idea of nominal-, ordinal-, and interval-level measurement. This seminal notion was introduced into social science in a classic article by S. S. Stevens in 1946. Stevens said that statistics like t

and r , because of certain assumptions that they made, require interval-level data, and this became an almost magical prescription.

Thirty-four years later, Gaito (1980) surveyed the (by then voluminous) mathematical statistics literature and found no support for the idea that measurement properties have anything to do with the selection of statistical procedures. Social scientists, said Gaito, confuse measurement (which focuses on the meaning of numbers) with statistics (which doesn't care about meaning at all) (p. 566). So, treating ordinal variables as if they were interval, for purposes of statistical analysis, is almost always a safe thing to do, especially with five or more ordinal categories (Boyle 1970; Labovitz 1971a).

The important thing is measurement, not statistics. As I pointed out in chapter 2, many concepts, such as gender, race, and class are much more subtle and complex than we give them credit for being. Instead of measuring them qualitatively (remember that assignment of something to a nominal category is a qualitative act of measurement), we ought to be thinking hard about how to measure them ordinally.

Emil Durkheim was an astute theorist. He noted that the division of labor became more complex as the complexity of social organization increased (Durkheim 1933 [1893]). But he, like other theorists of his day, divided the world of social organization into a series of dichotomous categories (*Gemeinschaft* vs. *Gesellschaft*, or mechanical vs. organic solidarity).

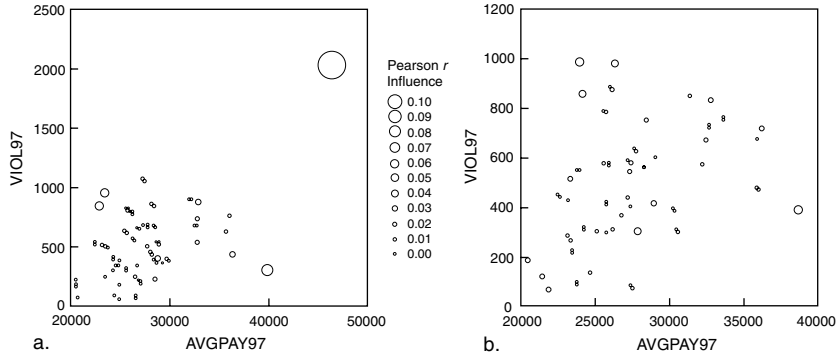
Today, social theorists want to know how degrees of differences in aspects of social organization (like the division of labor in society) are related to social complexity. This requires some hard thinking about how to measure these two variables with more subtlety. The meaning of the measurements is crucial.

Eliminating the Outliers

Another controversial practice in data analysis is called “eliminating the outliers,” which means removing extreme values from data analysis. If there are clear indications of measurement error (a person with a score of 600 on a 300-point test turns up in your sample), you can throw out the data that are in error. The problem comes when outliers (so-called freak cases) are eliminated just to smooth out data—to achieve better fits of regression lines to data. A single wealthy household might be ignored in calculating the average household income in a community on the theory that it's a “freak case.” But what if it isn't a freak case? What if it represents a small, but substantively significant proportion of cases in the community? Eliminating it only prevents the discovery of that fact.

Trivially, you can always achieve a perfect fit to a set of data if you reduce it to just two points. But is creating a good fit what you're after? Don't you

really want to understand what makes the data messy in the first place? In general, you cannot achieve understanding of messiness by cleaning things up. Still, as in all aspects of research, be ready to break this rule, too, when you think you'll learn something by doing so.



Violent crimes in the United States (for all states and Washington, D.C.), per 100,000 population, by average annual pay, 1997.

Violent crimes in the United States (for all states, without Washington, D.C.), per 100,000 population, by average annual pay, 1997.

Figure 20.9. The effect of outliers.

Figure 20.9a shows the relation between (1) the number of violent crimes per 100,000 people in each of the 50 U.S. states and Washington, D.C., in 1997 and (2) the average annual pay for people in the 50 states plus Washington, D.C. Figure 20.9b shows exactly the same thing, but without including the data from Washington, D.C. The correlation between the two variables *with* D.C. in the picture is .44. When we leave out the data for D.C., the correlation sinks to .19.

That's because the violent crime rate in Washington, D.C., was an appalling 2,024 per 100,000 population in 1997, and the average annual pay there was \$46,775. For the 50 states in the United States, the next highest violent crime rate, in Florida, was half that of D.C., and the next highest average pay, in Connecticut, was 17% lower than that in D.C.

The sizes of the dots in figures 20.9a and b indicate the influence each data point has on the correlation—that is, how much the correlation would change if you took each data point out of the calculation. The legend for this little trick is between the figures. That huge circle in the upper-right of figure 20.9a is Washington, D.C. To repeat: Pearson's *r* for the two variables in figure 20.9a is 0.442 and for figure 20.9b, without D.C., it's 0.190. Now, *that's* an outlier.

What to do with outliers? Try using the median to describe the central tendency rather than eliminating cases. Or report the results of your analysis with and without outliers, as I just did with figures 20.9a and b.

Tests of Significance

This is a hot topic in social science. Some years ago, researchers began arguing that statistical tests of significance are virtually useless (Rozeboom 1960; Labovitz 1971b; Carver 1978), and the drumbeat has continued ever since (Carver 1993; Cohen 1994).

I wouldn't go that far. Consider the hypothesis that the universe is expanding. As Wainer (1999) points out, being able to reject the null hypothesis at $p < .05$ would be quite a contribution. Will Dr. X be denied tenure a year from now? Lots of people, says Wainer, would be happy to know that they could reject the null hypothesis at, say, $p < .001$.

It's true that if you don't have a representative sample, then a test of statistical significance doesn't allow you to generalize beyond your particular sample of data. On the other hand, if you get significant results on a nonrandom sample, at least you can rule out the operation of random properties *in your sample* (Blalock 1979:239ff).

Use tests of significance but remember that they aren't magical and that the .01 and .05 levels of significance, while widely accepted tribal customs, are not sacred. They are simply conventions that have developed for convenience over the years. Greenwald et al. (1996:181–182) offer some useful advice about reporting p values.

1. In many situations, it's enough to use simple asterisks in your prose to indicate statistical significance. But if you need to report p values, then do so with an = sign, not with a > or < sign. If a p value is .042, don't report it as $p < .05$ ("the probability is less than .05"). Just report it as $p = .042$ and be done with it. Probability, just like the confidence we have in probabilistic results, is a continuous variable. Why cut out all that information with arbitrary cutoffs?
2. A single p value of .05 is an indicator of a relation, but is not convincing support for a hypothesis. By tradition, researchers almost never report probabilities that are greater than .05. Five *repeated results* of $p = .06$, or even .10, are more convincing than a single result of $p = .05$ that something's really going on.

THE BONFERRONI CORRECTION

If you want to be especially cautious in reporting correlations, you can apply a test known as the **Bonferroni correction**. There are two kinds of spu-

rious correlation. One happens when the correlation between two variables is caused by a third, independent, variable. Another occurs when two variables covary because of sheer accident. It happens all the time. In fact, it occurs with a known probability.

Take a random sample of, say, 30 variables in the world, measure them, and correlate all possible pairs of those variables. (There are $n(n-1)$ pairs of 30 anything.) According to the Bonferroni rule, you should get a correlation, significant at the $p = \leq .05$ level, in 5% of all cases—in 22 out of 435 pairs—just by chance. So, if you build a matrix with 435 pairs of correlations and find that 20 of them are significant at the .05 level or better, you can't be sure that this is not simply a random event.

Of course, you *always* have to explain why any two variables are correlated, since correlation, by itself, never implies cause and effect. But when you go fishing for significant correlations in a big matrix of them, and find fewer than you'd expect by chance alone, things are even tougher.

How tough? Pick a level of significance for reporting findings in your data—say, .05. If you have 30 variables in your analysis, and 435 tests of covariations in your matrix, divide .05 by $435 = .0001$. If you report these correlations as significant at the 5% level (the level you chose originally), then, according to the Bonferroni rule, your report will be valid (see Koopmans 1981; Kirk 1982). This is a very, very conservative test, but it will prevent you from making those dreaded Type I errors and reporting significant relations that aren't really there.

On the other hand, this will increase your chance of making Type II errors—rejecting some seemingly insignificant relations when they really *are* important. You might fail to show, for example, that certain types of exposure are related to contracting a particular disease, and this would have negative public health consequences. There's no free lunch.

Consider the study by Dressler (1980). He studied a sample of 40 people in the Caribbean island of St. Lucia, all of whom had high blood pressure. Dressler measured nine variables having to do with his respondents' ethnomedical beliefs and their compliance with a physician-prescribed treatment regimen. He reported the entire matrix of $(9 \times 8)/2 = 36$ correlations, 13 of which were significant at the 5% level or better.

Dressler might have expected $36 \times .05 = 1.8$ such correlations by chance. Three of the 13 correlations were significant at the .001 level. According to the Bonferroni rule, correlations at the $.05/36 = .0014$ level would be reportable at the .05 level as valid. Under the circumstances, however (13 significant correlations with only about two expected by chance), Dressler was quite justified in reporting all his findings.

I feel that if you are doing fieldwork, and using small data sets, you should

be comfortable with tests of significance at the $p < .10$ level, especially if you can repeat your finding in independent tests. Two or three repeated experiments that produce similar results in the same direction at the .10 level of significance are more convincing evidence of something going on than is one experiment that just barely pans out at the .05 level.

On the other hand, you can always find significant covariations in your data if you lower alpha (the level of significance) enough, so be careful. Remember, you're using statistics to get hints about things that are going on in your data. I cannot repeat often enough the rule that real analysis (building explanations and suggesting plausible mechanisms that make sense out of covariations) is what you do *after* you do statistics.

I also can't stress enough the difference between **statistical significance** and **practical significance**. If you have a large enough n (in the thousands), you will surely find significant statistical relations in your data. Each of those relations may account for a small amount of the variance in what you're interested in understanding, but that doesn't mean that the relations are of practical use.

Lots of research across the world shows a strong correlation between instruction about the dangers of unprotected sex and an increase in knowledge about HIV/AIDS, but not necessarily any changes in risky sexual behavior. Instruction in the use of contraceptives produces an increase in knowledge about reproductive health, but not necessarily a decrease in fertility.

Even when an increase in knowledge does produce desired changes in behavior, the practical significance of this information may be set aside by political realities. If you start early enough, teaching children about the dangers of smoking apparently produces a reduction in smoking. Suppose you come up with a curriculum that costs \$60,000 to implement in a school district of 2,000 students and that produces an aggregate reduction of 10% in smoking behavior and that the 10% is a statistically significant reduction. Will the school board shell out this money? The statistical level of significance in the results may play *some* role in the board's decision, but I'll bet that other things will weigh even more heavily in their deliberations.

STATISTICAL POWER

The **power** of a statistical test is the probability of *correctly accepting your research hypothesis*. If you're thinking: "You mean it's the probability of taking 'yes' for an answer?" then you're right on track. As you know, the traditional way to conduct research is to: (1) formulate a hypothesis; (2) turn the hypothesis around into a null hypothesis; and then (3) try as hard as we can to prove the null hypothesis.

This is *not* perverse. All of us positivists out here know that it's impossible to absolutely, positively, prove any hypothesis to be forever unfalsifiably true. So we do the next best thing. We try our very best to disprove our best ideas (our research hypotheses) and hope that we fail, leaving us with the right to say that our best guess is that we were right to begin with.

What this means in the real life of researchers is that statistical power is the probability of avoiding *both* Type I *and* Type II errors: rejecting the null hypothesis when it's really true (Type I error) or accepting a null hypothesis when it's really false (Type II error). This probability depends on two things: (1) the minimum size of the difference between two outcomes that you will accept as a *real* difference and (2) the size of the sample. So, to achieve a given amount of statistical power in any experiment or survey, you need to calculate the size of the sample required, given the minimum size of the difference between two outcomes—the effect size—that you will accept as a real difference (Kraemer and Thieman 1987; Cohen 1988).

This is a very important and subtle issue. Suppose you ask 100 men and 100 women, matched for socioeconomic status, race, and religion, to take the Attitudes Toward Women Scale (AWS). The null hypothesis is that there is no difference between the mean scores of the men and the mean scores of the women on this scale. How big a difference do you need between the mean of the men and the mean of the women on this scale to reject the null hypothesis and conclude that, in fact, the difference is real—that men and women really differ on their attitudes toward women as expressed in the AWS?

The answer depends on the power of the test of the difference in the means. Suppose you analyze the difference between the two means with a *t*-test, and suppose that the test is significant at the .05 level. Statistical power is the probability that you are wrong to report this result as an indicator that you can reject the null hypothesis.

The result, at the $p = .05$ level, indicates that the difference you detected between the mean for the men and the mean for the women would be expected to occur by chance fewer than five times in 100 runs of the same experiment. It does *not* indicate that you are $1 - p$, or 95% confident that you have correctly rejected the null hypothesis. The power of the finding of $p = .05$ depends on the size of the sample and on the size of the difference that you expected to find before you did the study.

In the case of the AWS, there are 30 years of data available. These data make it easy to say how big a difference you expect to find if the men and women in your sample are really different in their responses to the AWS. Many surveys, especially those done in foreign fieldwork, are done without this kind of information available. You can offer a theory to explain the results from one experiment or survey. But you can't turn around and use those same

data to *test* your theory. As replications accumulate for questions of importance in the social sciences, the question of statistical power becomes more and more important.

So, what's the right amount of statistical power to shoot for? Cohen (1992) recommends that researchers plan their work—that is, set the effect size they recognize as important, set the level of statistical significance they want to achieve (.05, for example, or .01), and calculate the sample size—in order to achieve a power of .80.

A power value of .80 would be an 80% chance of recognizing that our original hypothesis is really true and a 20% chance of rejecting our hypothesis when it's really true. If you shoot for a power level of much lower than .80, says Cohen, you run too high a risk of making a Type II error.

On the other hand, power ratings much higher than .80 might require such large *n*'s that researchers couldn't afford them (*ibid.*:156). If you want 90% power for a .01 (1%) two-tailed test of, say, the difference between two Pearson's *r*'s, then, you'd need 364 participants (respondents, subjects) to detect a difference of .20 between the scores of the two groups. If you were willing to settle for 80% power and a .05 (5%) two-tailed test, then the number of participants drops to 192. (To find the sample size needed for any given level of power, see Kraemer and Theimann 1987:105–112.)

The Shotgun Approach

A closely related issue concerns “shotgunning.” This involves constructing a correlation matrix of all combinations of variables in a study, and then relying on tests of significance to reach substantive conclusions.

Kunitz et al. (1981) studied the determinants of hospital utilization and surgery in 18 communities on the Navajo Indian Reservation during the 1970s. They measured 21 variables in each community, including 17 independent variables (the average education of adults, the percentage of men and women who worked full-time, the average age of men and women, the percentage of income from welfare, the percentage of homes that had bathrooms, the percentage of families living in traditional hogans, etc.) and four dependent variables (the rate of hospital use and the rates for the three most common types of surgery). Table 20.20 shows the correlation matrix of all 21 variables in this study.

There are $n(n - 1)/2$ pairs in any list of items, so, for a symmetric matrix of 21 items there are 210 possible correlations. Kunitz et al. point out in the footnote to their matrix that, for $n = 18$, the 0.05 level of probability corresponds to $r = 0.46$ and the 0.01 level corresponds to $r = 0.56$. By the Bonferonni correction, they could have expected:

TABLE 20.20
 Correlation Matrix of All 21 Variables in Kunitz et al.'s Study of Hospital Use on the Navajo Reservation

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1 Near hospital																					
2 Near surgery	.67																				
3 Wage work	-.24	-.09																			
4 Welfare	.52	.46	-.54																		
5 Education of men	-.42	-.43	.73	-.49																	
6 Education of women	.01	-.21	.67	-.32	.81																
7 Hogans	.07	.37	-.26	.72	-.28	-.40															
8 Bathrooms	-.44	-.57	.63	-.63	.70	.64	-.47														
9 Household size	.01	-.16	-.48	.24	-.34	-.07	.04	.12													
10 Working women	-.35	-.36	.68	-.65	.65	.57	-.48	.62	-.22												
11 Working men	-.24	-.18	.73	-.37	.73	.63	-.16	.41	-.45	.45											
12 Vehicles	.34	-.08	.40	-.17	.22	.52	-.53	.29	-.06	.31	.40										
13 Median income	-.45	-.51	.66	-.60	.67	.60	-.38	.79	-.07	.83	.48	.27									
14 Per capita income	-.25	-.20	.68	-.46	.61	.40	-.15	.29	-.68	.48	.51	.30	.47								
15 Age of women	-.23	-.03	-.47	-.15	-.36	.66	-.02	-.46	-.30	-.37	-.32	-.56	-.43	-.04							
16 Age of men	.28	.15	-.77	.31	-.60	-.55	.01	-.67	.06	-.51	-.68	-.35	-.63	-.45	.60						
17 Age of patients	.13	.47	.13	-.14	-.21	.15	-.18	-.22	-.36	-.04	-.26	-.15	-.26	-.02	.29	.32					
18 Hysterectomies	-.41	-.48	.62	-.55	.57	.46	-.23	.75	-.09	.45	.46	.39	.62	.59	-.37	-.62	0.28				
19 Appendectomies	-.31	-.40	.32	-.14	.44	.33	.16	.51	.13	.07	.27	.01	.42	.35	-.31	-.49	-.56	-.62			
20 Cholelsectomies	-.23	-.70	.15	-.43	.35	.34	-.46	.68	.22	.19	.12	.31	.40	.14	-.17	-.17	-.35	.70	.49		
21 Hospital rate	-.49	.24	.02	-.26	.02	-.25	-.09	.16	-.19	-.04	-.34	-.33	.01	.22	.43	0.16	.45	.17	.10	.18	

N = 18 0.46, $p = .05$ 0.56, $p = 0.01$

SOURCE: Reprinted from *Social Science and Medicine* 15B, S. J. Kunitz et al., "Determinants of Hospital Utilization and Surgery on the Navajo Indian Reservation, 1972-1978," pp. 71-79, © 1981, with permission from Elsevier.

$$[(21 \times 20)/2](.05) = 10.5$$

correlations significant at the 0.05 level and

$$[(21 \times 20)/2](.01) = 2.1$$

correlations significant at the 0.01 level by chance. There are 73 correlations significant at the 0.05 level in table 20.20, and 42 of those correlations are significant at the 0.01 level.

Kunitz et al. examined these correlations and were struck by the strong association of the hysterectomy rate to all the variables that appear to measure acculturation. I'm struck by it, too. This interesting finding was not the result of deduction and testing; it was the result of shotgunning. The finding is not proof of anything, of course, but it sure seems like a strong clue to me. I'd want to follow this up with research on how acculturation affects the kind of medical care that women receive and whether all those hysterectomies are necessary.

The Problem with the Shotgun Approach

The problem with shotgunning is that you might be fooled into thinking that *statistically* significant correlations are also *substantively* significant. This is a real danger, and it should not be minimized (Labovitz 1972). It results from two problems.

1. You have to be very careful about choosing a statistical measure of association, depending on how the variables were measured in the first place. A significant correlation in a matrix may be an artifact of the statistical technique used and not be of any substantive importance. Running a big correlation matrix of all your variables may produce some statistically significant results that would be insignificant if the proper test had been applied.
2. There is a known probability that any correlation in a matrix might be the result of chance. The number of expected significant correlations in a matrix is equal to the level of significance you choose, times the number of variables. If you are looking for covariations that are significant at the 5% level, then you only need 20 tests of covariation to find one such covariation by chance. If you are looking for covariations that are significant at the 1% level, you should expect to find one, by chance, in every 100 tries. In a matrix of 100 variables with 4,950 correlations, you might find around 50 significant correlations at the 1% level by chance.

This does not mean that 50 correlations at the 1% level in such a matrix *are* the result of chance. They just *might* be. There could be 100 or more signifi-

cant correlations in a symmetric matrix of 100 variables. If 50 of them (4,950/100) might be the result of chance, then how can you decide which 50 they are? You can't. You can never know for sure whether any particular correlation is the result of chance. You simply have to be careful in your interpretation of every correlation in a matrix.

Use the shotgun. Be as cavalier as you can in looking for statistically significant covariations, but be very conservative in interpreting their substantive importance. Correlations are hints to you that something is going on between two variables. Just keep in mind that the leap from correlation to cause is often across a wide chasm.

If you look at table 20.18 again, you can see just how risky things can be. A correlation of .60 is significant at the 1% level of confidence with a sample as small as 30. Notice, however, that the correlation in the population is 99% certain to fall between .20 and .83, which is a pretty wide spread. You wouldn't want to build too big a theory around a correlation that just might be down around the .20 level, accounting for just 4% of the variance in what you're interested in!

Remember these rules:

1. Not all significant findings at the 5% level of confidence are equally important. A very weak correlation of .10 in a sample of a million people would be statistically significant, even if it were substantively trivial. By contrast, in small samples, substantively important relations may show up as statistically insignificant.
2. Don't settle for just one correlation that supports a pet theory; insist on several, and be on the lookout for artifactual correlations.

Fifty years ago, before statistical packages were available, it was a real pain to run any statistical tests. It made a lot of sense to think hard about which of the thousands of possible tests one really wanted to run by hand on an adding machine.

Computers have eliminated the drudge work in data analysis, but they haven't eliminated the need to think critically about your results. If anything, computers—especially those little ones that fit in your backpack and have full-featured statistics packages that do everything from χ^2 to factor analysis—have made it more important than ever to be self-conscious about the interpretation of statistical findings. But if you *are* self-conscious about this issue, and dedicated to thinking critically about your data, then I believe you should take full advantage of the power of the computer to produce a mountain of correlational hints that you can follow up.

Finally, by all means, use your intuition in interpreting correlations; common sense and your personal experience in research are powerful tools for

data analysis. If you find a correlation between the number of times that men have been arrested for drug dealing and the number of younger siblings they have, you'd suspect that this correlation might be just a chance artifact.

On the other hand, maybe it isn't. There is just as much danger in relying slavishly on personal intuition and common sense as there is in placing ultimate faith in computers. What appears silly to you may, in fact, be an important signal in your data. The world is filled with self-evident truths that aren't true, and self-evident falsehoods that aren't false. The role of science, based on solid technique and the application of intuition, is to sort those things out.

21



Multivariate Analysis

Most of the really interesting dependent variables in the social world—things like personality type, amount of risk-taking behavior, level of wealth accumulation, attitudes toward women or men—appear to be caused by a large number of independent variables, all of which are dependent variables themselves. The goal of multivariate analysis is to test hypotheses about *how* variables are related, based on a theory of causation. This is called **causal modeling**.

Multivariate analysis involves an array of statistical procedures—things like multiple regression, partial regression, factor analysis, multidimensional scaling, and so on. I'll introduce you to the conceptual basis of some of these methods here. I hope that this will give you an idea of the range of tools available and enough information so you can read and understand research articles in which these techniques are used. I also hope that this will arouse your curiosity enough so that you'll study these methods in more advanced classes. This is the fun part.

Elaboration: Controlling for Independent Variables

We begin with the **elaboration method** developed by Paul Lazarsfeld and his colleagues (1972) for analyzing data from surveys (see especially Rosenberg 1968). This method involves teasing out the complexities in a bivariate relation by **controlling for** the effects of a third (antecedent or intervening) variable. It was popular in sociology until the mid-1970s, but fell out of favor when computers became common and made lighter work of multivariate analysis.

I still think that the method is excellent for use in anthropological fieldwork. This is not because elaboration analysis can be done without a computer. It *can* be done without a computer, but you know how I feel about that: Once you understand how any statistical method works, you should *only* do it by computer. The elaboration method is wonderful for field researchers because it keeps you close to your data.

It's going to take you a while to get through the next half-dozen pages on the elaboration method. The writing is clear and there's nothing more complicated than a chi-square, so they're not tough going. They're just plain tedious. But bear with me. Eventually, you'll give a computer a list of independent variables, specify a dependent variable, and let the machine do the heavy lifting. But the next few pages will give you an appreciation of what a multivariate analysis does. They will also give you the skills you need to conduct a multivariate analysis in the field, while your thoughts are fresh and you still have time to collect any data you find you need. So be patient, pay close attention to the tables, and stay with it.

Building Tables

Suppose you are working in Peru and you suspect that Indians who move to Lima are no better off than Indians who remain in the villages. The Indians claim that they are seeking better jobs and better opportunities for their children, but you think that they are not getting what they came to the city to find. You conduct a survey of 250 village residents from a particular region and 250 migrants who have gone to the city from the same region. Table 21.1

TABLE 21.1
Wealth by Residence for a Sample of 500 Peruvian Indians (Hypothetical Data)

<i>Wealth</i>	<i>Rural</i>	<i>Urban</i>	<i>Row totals</i>
Not poor	84	91	175
Poor	166	159	325
Column totals	250	250	500
$\chi^2 = .56$ ns (not significant)			

shows the relation between residence and accumulated wealth status for your sample.

Chi-square for this table is not significant. Assuming that you have measured wealth status using an appropriate index for both the urban and village

environments, residence appears to make no difference in wealth accumulation among these informants.

Table 21.2, however, shows that after 5 years or more in the city, 68% (28/41) were no longer poor by local standards. In fact, the odds of emerging from

TABLE 21.2
Wealth by Time in the City for Urban Migrants

<i>Wealth</i>	<i><1 year</i>	<i>1–3 years</i>	<i>3–5 years</i>	<i>>5 years</i>	<i>Row totals</i>
Not poor	6	10	12	28	56
Poor	77	60	44	13	194
Column totals	83	70	56	41	250

$\chi^2 = 63.35, p < .001$

poverty increase steadily over time. There is no guarantee, of course, but the Indians' perception that urban migration works in their favor is substantially correct from these data. Thus, controlling for time allows the significant bivariate relation between wealth and urban residence to emerge.

While time is an important factor in understanding *what's* going on, it doesn't tell us anything about *how* the process works. We know from many studies that education is related positively to wealth. Table 21.3 shows the

TABLE 21.3
Wealth by Education for the Entire Sample

<i>Wealth</i>	<i>Did not complete 8th grade</i>	<i>Completed 8th grade</i>	<i>Row totals</i>
Not poor	62	113	175
Poor	179	146	325
Column totals	241	259	500

$\chi^2 = 17.59, p < .001, OR = 2.24$

breakdown for the entire sample of 500, including the rural stay-at-homes and the urban migrants. Chi-square is highly significant, and the odds ratio for this table is $OR = 2.24$. The odds of being poor for those who haven't completed 8 years of school are more than double those of people who have completed the eighth grade.

Table 21.4 breaks down education by residence. Chi-square for this table is also highly significant, and the odds ratio is 2.62. In other words, urbanites are more than two-and-a-half times as likely to complete *secundaria*, or the eighth grade. This, it appears, leads to greater wealth.

We can test this hypothesis by producing a single cross-tab of wealth by

TABLE 21.4
Education by Residence

<i>Education</i>	<i>Rural</i>	<i>Urban</i>	<i>Row totals</i>
Completed 8th grade	100	159	259
Did not complete 8th grade	150	91	241
Column totals	250	250	500

$\chi^2 = 27.88, p < .001, OR = 2.62$

education, controlling for residence. This is done in table 21.5, which really consists of two separate tables, each of which can be analyzed statistically. (Place the control variables above the independent variable when constructing multivariate tables.)

TABLE 21.5
Wealth by Education, Controlling for Residence

	<i>Rural</i>			<i>Urban</i>		
	$\geq 8th\ grade$	$< 8th\ grade$	<i>Row totals</i>	$\geq 8th\ grade$	$< 8th\ grade$	<i>Row totals</i>
Not poor	50	34	84	63	28	91
Poor	50	116	166	96	63	159
Column totals	100	150	250	159	91	250

$\chi^2 = 20.09, p < .001, OR = 3.41$ $\chi^2 = 1.96\ ns\ (not\ significant)$

Things are a bit more complex than we imagined at first. Education makes a difference in the odds of escaping poverty, but only among rural people. What's going on here? To find out, we continue to elaborate the analysis, looking at other variables and how they may be magnifying or suppressing relationships.

As you add variables, of course, the number of tables required goes up, *as does the required sample size*. Adding a third variable, residence, to the analysis of wealth by education, requires two additional tables: residence by wealth and residence by education. Adding family size to the model, we need *three* additional tables. Tables 21.6, 21.7, and 21.8 show the breakdown for family size by education, wealth by family size, and family size by residence.

In table 21.6, we see that people with more education tend to have smaller families. In table 21.7, we see that poor families tend to have more children. And in table 21.8, we see that rural people tend to have more children. It appears from these tables that economic status may be related more strongly to family size than to education or to residence.

TABLE 21.6
Family Size by Education

	$\geq 8th\ grade$	$< 8th\ grade$	<i>Row totals</i>
≤ 3 children	170	129	299
> 3 children	89	112	201
Column totals	259	241	500

$\chi^2 = 7.62, p < .01, OR = 1.65$

TABLE 21.7
Family Size by Wealth

	$> 3\ children$	$\leq 3\ children$	<i>Row totals</i>
Not poor	84	91	175
Poor	215	110	325
Column totals	299	201	500

$\chi^2 = 15.59, p < .001, OR = 2.12$

TABLE 21.8
Family Size by Residence

	<i>Rural</i>	<i>Urban</i>	<i>Row totals</i>
> 3 children	167	132	299
≤ 3 children	83	118	201
Column totals	250	250	500

$\chi^2 = 10.19, p < .001, OR = 1.80$

To disentangle things, we look at the original relationship between wealth and residence, controlling for family size. Table 21.9 shows that when we do that, the effect of residence on economic status remains insignificant for rural people, but it makes a big difference for urban residents. When we look at the relationship between wealth and education in table 21.10, we see that the influence of education on wealth is insignificant for large families, but is highly significant for small families.

To get the full picture, we now produce table 21.11, which shows the bivariate relationship between wealth status and education, now controlling for *both* family size and residence simultaneously.

TABLE 21.9
Wealth by Residence, Controlling for Family Size

	<i>Rural</i>			<i>Urban</i>		
	<i>Number of children</i>		<i>Row totals</i>	<i>Number of children</i>		<i>Row totals</i>
	<i>>3</i>	<i>≤3</i>		<i>>3</i>	<i>≤3</i>	
Not poor	54	30	84	30	61	91
Poor	113	53	166	102	57	159
Column totals	167	83	250	132	117	250
$\chi^2 = .36$ ns (not significant)			$\chi^2 = 22.58$, OR = 3.64			

TABLE 21.10
Wealth by Education, Controlling for Family Size

	<i>>3 children</i>			<i>≤3 children</i>		
	<i>Education</i>		<i>Row totals</i>	<i>Education</i>		<i>Row totals</i>
	<i>≥8 years</i>	<i><8 years</i>		<i>≥8 years</i>	<i><8 years</i>	
Not poor	54	30	84	59	32	91
Poor	116	99	215	30	80	110
Column totals	170	129	299	89	112	201
$\chi^2 = 2.63$ ns (not significant)			$\chi^2 = 24.48$, OR = 4.92			

Reading a Complex Table

Reading across table 20.11, we see that among urban families with at least an eighth-grade education and three or fewer children, 43 out of 59, or 73%, are above the poverty line. Among urban families with at least an eighth-grade education and more than three children, only 20% (20/100) are above the poverty line. Among rural families with at least an eighth-grade education and with three or fewer children, 53% are above the poverty line. Among rural people with at least an eighth-grade education and more than three children, 49% are above the poverty line.

In other words, for rural people, education appears to be the key to rising above poverty. So long as they increase their education, they are about as likely (49% versus 53%) to increase their economic status, whether or not they limit natality.

For urban migrants, the picture is different. We saw from table 21.2 that the

TABLE 21.11
Wealth by Education, Controlling for Family Size and Residence

	<i>Rural</i>				<i>Urban</i>				<i>Row totals</i>
	<i>> 3 children</i>	<i><8th grade</i>	<i>≥8th grade</i>	<i>≤3 children</i>	<i>> 3 children</i>	<i><8th grade</i>	<i>≥8th grade</i>	<i>≤3 children</i>	
Not poor	34	20	16	14	20	10	43	18	91
Poor	36	77	14	39	80	22	16	41	159
Column totals	70	97	30	53	100	32	59	59	250

longer the urban migrants remained in the city, the greater the likelihood that they would rise above poverty. Now we know that unless they lower their natality *and* increase their education, the poverty-stricken villagers in our sample are probably better off staying home and not migrating to the city. If the urban migrants in our sample (all of whom started out as poor villagers) fail to limit natality, they lose the advantage that education would otherwise bring them. Rural people keep this advantage, irrespective of family size.

Explaining this finding, of course, is up to you. That's what theory is all about. A causal connection between variables requires a mechanism that explains how things work. (See chapter 2 if you need to go over the issue of covariation and causality.) In this instance, we might conjecture that rural people have lower overall expenses, especially if they own their own land and homes. They usually have extended families that cut down the cost of child care and that provide no-interest loans during emergencies. They grow much of their own food. And having more children may help them farm more land and cut down on expenses.

Urban people get more education, and this gets them better paying jobs. But if they have many mouths to feed, and if they have to pay rent, and if they lack the financial support of kin close by, then these factors may vitiate any advantage their education might otherwise bring.

Taking the Elaboration Another Step

We can look for clues that support or challenge our theory by elaborating the model still further, this time using family size as the dependent variable. Table 21.12 shows the result of cross-tabulating family size by education, controlling for residence.

Chi-square for the left half of this table is insignificant, but for the right half

TABLE 21.12
Family Size by Education, Controlling for Residence

	<i>Rural</i>			<i>Urban</i>		
	$\geq 8th$ <i>grade</i>	$< 8th$ <i>grade</i>	<i>Row</i> <i>totals</i>	$\geq 8th$ <i>grade</i>	$< 8th$ <i>grade</i>	<i>Row</i> <i>totals</i>
>3 children	70	97	167	100	32	132
≤ 3 children	30	53	83	59	59	118
Column totals	100	150	250	159	91	250
$\chi^2 = .77$ ns (not significant)			$\chi^2 = 17.86$, $p < .001$, OR = 3.13			

it is highly significant. Rural informants with less than an eighth-grade education are more likely than urban informants with less than an eighth-grade education to have more than three children (65% versus 35%). Among rural informants, in fact, level of education has little or no effect on family size (70% of those with higher education have large families versus 65% of those with lower education).

Among urban informants, education is strongly associated with family size. More-educated urban informants are more likely than less-educated informants to have more than three children, according to these data. This throws new light on the subject, and begs to be explained. We know that higher education without small families does not produce an increase in economic status for these poor migrants. We know, too, that most people, whether urban or rural, keep having large families, although large families are less prevalent among urbanites than among rural residents (132 out of 250 versus 167 out of 250).

To understand this case still further, consider table 21.13, which cross-tabulates family size by wealth, controlling for both education and residence. This

TABLE 21.13
Family Size by Wealth, Controlling for Education and Residence

	<i>Rural</i>					<i>Urban</i>				
	$\geq 8th\ grade$		$< 8th\ grade$			$\geq 8th\ grade$		$< 8th\ grade$		
	<i>Not Poor</i>	<i>poor</i>	<i>Not Poor</i>	<i>poor</i>	<i>Row totals</i>	<i>Not Poor</i>	<i>poor</i>	<i>Not Poor</i>	<i>poor</i>	<i>Row totals</i>
>3 children	34	36	20	77	167	20	80	10	22	132
≤ 3 children	16	14	14	39	83	43	16	18	41	118
Column totals	50	50	34	116	250	63	96	28	63	250

table shows that neither wealth nor education influences family size among rural informants. For urban residents, however, the story is quite different. As expected, those urban informants who have both increased their education and increased their wealth have small families.

Go through table 21.13 carefully and make the appropriate comparisons across the rows and between the two halves. Compare also the results of this table with those of table 21.11, in which wealth status was the dependent variable.

From these tables, we can now hazard a guess about how these variables interact. We can draw a **conceptual model** of the whole process we've been looking at. It's in figure 21.1.

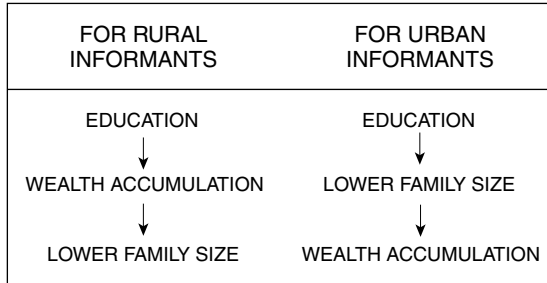


Figure 21.1. Model of how wealth, education, and family size interact in urban and rural environments for informants in tables 21.11 and 21.13.

Most people in our sample are poor. Sixty-six percent of rural informants (166/250) and 64% of urban informants (159/250) are below the poverty line by our measurements. Among rural informants, education provides an edge in the struggle against poverty, irrespective of family size, but for urban migrants, education only provides an edge in the context of lowered family size.

Among those who remain in the villages, then, education may lead either to accumulation of wealth through better job opportunities, or it may have no effect. The chances are better, though, that it leads to wealth. Once this occurs, it leads to control of fertility. Among urban informants, education leads either to control of natality or not. If not, then education has practically no effect on the economic status of poor migrants. If it does lead to lowered natality, then it may lead, over time, to a favorable change in economic status.

We can check this model by going back to our data on wealth status by number of years in the city to see if those migrants who are economically successful over time have both increased their education *and* lowered their natality. Plausible assumptions about time ordering of variables are crucial in building causal models. Knowing, for example, that wealthy villagers never move to the city rules out some alternative explanations for the data presented here.

You get the picture. The elaboration method can produce subtle results, but it is quite straightforward to use and depends only on your imagination, on simple arithmetic (percentages), and on basic bivariate statistics.

Partial Correlation

Of course, I haven't proven anything by all this laying out of tables. Don't misunderstand me. Elaboration tables are a great start—they test whether your

ideas about some antecedent or intervening variables are plausible by showing what *could* be going on—but they don’t tell you *how things work* or *how much* those antecedent or intervening variables are contributing to a correlation you want to understand. For that, we need something a bit more . . . well, elaborate. Partial correlation is a direct way to control for the effects of a third (or fourth or fifth . . .) variable on a relationship between two variables.

Here’s an interesting case. Across the 50 states in the United States, there is a stunning correlation ($r = .778$) between the percentage of live births to teenage mothers (15–19 years of age) and the number of motor vehicle deaths per hundred million miles driven. States that have a high rate of road carnage have a high rate of births to teenagers, and vice versa.

This one’s a real puzzle. Obviously, there’s no *direct* relation between these two variables. There’s no way that the volume of highway slaughter causes the number of teenage mothers (or vice versa), so we look for something that might cause both of them.

I have a hunch that these two variables are correlated because they are both the consequence of the fact that certain regions of the country are poorer than others. I know from my own experience, and from having read a lot of research reports, that the western and southern states are poorer, overall, than are the industrial and farming states of the Northeast and the Midwest. My hunch is that poorer states will have fewer miles of paved road per million people, poorer roads overall, and older vehicles. All this might lead to more deaths per miles driven.

Table 21.14 shows the **zero-order correlation** among three variables: motor vehicle deaths per hundred million miles driven (it’s labeled MVD in

TABLE 21.14
Correlation Matrix for Three Variables

	<i>Variable 1</i> <i>MVD</i>	<i>Variable 2</i> <i>TEENBIRTH</i>	<i>Variable 3</i> <i>INCOME</i>
MVD	1.00		
TEENBIRTH	.778	1.00	
INCOME	-.662	-.700	1.00

SOURCE: MVD for 1995, Table 1018, *Statistical Abstract of the United States* (1997). TEENBIRTH for 1996, Table 98, *Statistical Abstract of the United States* (1997). INCOME for 1996, Table 706, *Statistical Abstract of the United States* (1997).

table 21.14); the percentage of live births to young women 15–19 years of age (TEENBIRTH); and average personal income (INCOME). Zero-order correlations do not take into account the influence of other variables.

We can use the formula for partial correlation to test directly what effect, if

any, income has on the correlation between TEENBIRTH and MVD. The formula for partial correlation is:

$$r_{12\cdot3} = \frac{r_{12} - [r_{32} r_{13}]}{\left[\sqrt{1 - r_{32}^2} \right] \left[\sqrt{1 - r_{13}^2} \right]} \quad \text{Formula 21.1}$$

where $r_{12\cdot3}$ means “the correlation between variable 1 (MVD) and variable 2 (TEENBIRTH), *controlling* for variable 3 (INCOME).”

Table 21.15 shows the calculation of the partial correlations for the entries in table 21.14.

TABLE 21.15
Calculating the Partial Correlations for the Entries in Table 21.14

Pairs	Pearson's r	r^2	$1 - r^2$	$\sqrt{1 - r^2}$
r_{12} (MVD and TEENBIRTH)	.778	0.60528	0.39472	0.628264
r_{13} (MVD and INCOME)	-.662	0.43824	0.56176	0.749504
r_{32} (TEENBIRTH and INCOME)	-.700	0.49	0.51	0.714143

So, the partial correlation between MVD and TEENBIRTH, controlling for INCOME is:

$$r_{12\cdot3} = \frac{.778 - [(-.700)(-.662)]}{(.714143)(.749504)} = \frac{.3146}{.535253} = .5877594$$

which we can round off to .59. (Remember, we have to use a *lot* of decimal places during the calculations in order to keep the rounding error in check. When we get through with the calculations we can round off to two or three decimal places.) In other words, when we **partial out** the effect of income, the correlation between MVD and TEENBIRTH drops from about .78 to about .59. That's because income is correlated with motor vehicle deaths ($r = -.662$) and with teenage births ($r = .49$).

If a partial correlation of .59 between the rate of motor vehicle deaths and the rate of teenage births still seems high, then perhaps other variables are at work. You can “partial out” the effects of two or more variables at once, but as you take on more variables, the formula naturally gets more complicated.

A simple correlation is referred to as a **zero-order correlation**. Formula 21.1 is for a **first-order correlation**. The formula for a **second-order correlation** (controlling for two variables at the same time) is:

$$r_{12\cdot34} = \frac{r_{12\cdot3} - [r_{14\cdot3} r_{24\cdot3}]}{\left[\sqrt{1 - r_{14\cdot3}^2} \right] \left[\sqrt{1 - r_{24\cdot3}^2} \right]} \quad \text{Formula 21.2}$$

You can also work out and test a model of how several independent variables influence a dependent variable all at once. This is the task of multiple regression, which we'll take up next. (For more on partial correlation, see Gujarati 2003.)

Multiple Regression

Partial correlation tells us how much a third (or fourth . . .) variable contributes to the relation between two variables. **Multiple regression** puts all the information about a series of variables together into a single equation that takes account of the interrelationships among independent variables. The result of multiple regression is a statistic called **multiple-R**, which is the combined correlation of a set of independent variables with the dependent variable, taking into account the fact that each of the independent variables might be correlated with each of the *other* independent variables.

What's really interesting is R^2 . Recall from chapter 20 that r^2 —the square of the Pearson product moment correlation coefficient—is the amount of variance in the dependent variable accounted for by the independent variable in a simple regression. R^2 , or **multiple-R squared**, is the amount of variance in the dependent variable accounted for by two or more independent variables simultaneously.

Now, if the predictors of a dependent variable were all uncorrelated with each other, we could just add together the pieces of the variance in the dependent variable accounted for by each of the independent variables. That is, it would be nice if $R^2 = r_1^2 + r_2^2 + r_3^2 . . .$

It's a real nuisance, but independent variables *are* correlated with one another. (This interdependence among independent variables is called **multicollinearity**, which we'll discuss in depth later.) What we need is a method for figuring out how much variance in a dependent variable is accounted for by a series of independent variables after taking into account all of the overlap in variances accounted for across the independent variables. That's what multiple regression does.

The Multiple Regression Equation

We covered basic regression in chapter 20, but just to bring you back up to speed, remember that in simple regression we use an equation that expresses how an independent variable is related to a dependent variable. On the left-hand side of the equation, we have the unknown score for y , the dependent variable. On the right-hand side, we have the y -intercept, called a . It's the

score for y if the dependent variable were zero. We have another coefficient, called b , that tells by *how much* to multiply the score on the independent variable for each unit change in that variable.

The general form of the equation (from chapter 20, formula 20.19) is:

$$y = a + bx$$

which means that the dependent variable, y , equals some constant plus another constant times the independent variable, x . So, for example, a regression equation like:

$$\text{Starting Annual Income} = \$22,000 + (\$4,000 \times \text{Years of College})$$

predicts that, on average, people with a high school education will start out earning \$22,000 a year; people with a year of college will earn \$26,000; and so on. A person with nine years of university education (say, someone who has a Ph.D.) would be predicted to start at \$58,000:

$$\text{Starting Annual Income} = \$22,000 + (\$4,000 \times 9) = \$58,000$$

Now suppose that the average starting salary for someone who has a Ph.D. is \$65,000. Several things could account for the discrepancy between our prediction and the reality. Sampling problems, of course, could be the culprit. Or it could be that there is just a lot of variability in starting salaries of people who have the Ph.D. English teachers who go to work in small, liberal arts colleges might start at \$40,000, while people who have a Ph.D. in finance and who go to work for major brokerage companies might start at \$150,000.

No amount of fixing the sample will do anything to get rid of the variance of starting salaries. In fact, the better the sample, the better it will reflect the enormous variance in those salaries.

In simple regression, if starting salary and years of education are related variables, we want to know “How accurately can we predict a person’s starting salary if we know how many years of education they have beyond high school?” In multiple regression, we build more complex equations that tell us how much each of *several* independent variables contributes to predicting the score of a single dependent variable.

A typical question for a multiple regression analysis might be “How well can we predict a person’s starting salary if we know how many years of college they have, *and* their major, *and* their gender, *and* their age, *and* their ethnic background?” Each of those independent variables contributes something to predicting a person’s starting salary after high school.

The regression equation for two independent variables, called x_1 and x_2 , and one dependent variable, called y , is:

$$y = a + b_1x_1 + b_2x_2 \qquad \textbf{Formula 21.3}$$

which means that we need to find a separate constant—one called b_1 and one called b_2 —by which to multiply each of the two independent variables. The general formula for multiple regression is:

$$y = a + b_1x_1 + b_2x_2 + \dots + b_nx_n \quad \text{Formula 21.4}$$

Recall that simple regression yields a PRE measure, r^2 . It tells you how much better you can predict a series of measures of a dependent variable than you could by just guessing the mean for every measurement. Multiple regression is also a PRE measure. It, too, tells you how much better you can predict measures of a dependent variable than you could if you guessed the mean—but using all the information available in a series of independent variables.

The key to regression are those ***b* coefficients**, or **weights**, in Formula 21.4. We want coefficients that, when multiplied by the independent variables, produce the best possible prediction of the dependent variable—that is, we want predictions that result in the smallest possible **residuals**. Those coefficients, by the way, are not existential constants. Remember that samples yield *statistics*, which only *estimate* parameters. Your statistics change with every sample you take and with the number of independent variables in the equation.

The MVD-TEENBIRTH Puzzle

Let's try to solve the puzzle of the relation between teenage births and the rate of motor vehicle deaths. We can take a stab at this using multiple regression by trying to predict the rate of teenage births *without* the data from motor vehicle deaths.

From our previous analyses with elaboration tables and with partial correlation, we already had an idea that income might have something to do with the rate of teen births. We know from the literature (Handwerker 1998) that poverty is associated with violence and with teenage pregnancy, and that this is true across ethnic groups, so I've added a variable on violent crimes that I think might be a proxy for the amount of violence against persons in each of the states.

Table 21.16 shows the correlation matrix for four variables: TEENBIRTH (the percentage of births to teenagers in the 50 U.S. states during 1996); INCOME (the mean per capita income for each of the 50 states during 1996); VIOLRATE (the rate of violent crime—rape, murder, assault, and robbery—per 100,000 population in the 50 states in 1995); and MVD (the number of motor vehicle deaths per 100 million vehicle miles in each of the 50 states in 1995).

We see right away that the mean per capita income predicts the rate of births to teenagers ($r = -.700$) almost as well as does the rate of motor vehicle

TABLE 21.16
Correlation Matrix for Variables Associated with the Percentage of
Teenage Births in the United States

	<i>TEENBIRTH</i>	<i>INCOME</i>	<i>VIOLRATE</i>	<i>MVD</i>
TEENBIRTH	1.00			
INCOME	-.700	1.00		
VIOLRATE	.340	.190	1.00	
MVD	.778	-.662	.245	1.00

deaths ($r = -.778$), and that mean income *also* predicts the rate of motor vehicle deaths rather well ($r = -.662$).

This is a clue about what might be going on: An antecedent variable, the level of income, might be responsible for the rate of motor vehicle deaths *and* the rate of teenage births. (By the way, did you notice that the strong correlations above were *negative*? The greater the mean income in the state, the lower the rate of teenage births and the lower the rate of motor vehicle deaths. Remember, correlations can vary from -1.0 to $+1.0$ and the strength of the correlation has nothing to do with its direction.)

And one more thing: The rate of violent crimes against people is moderately correlated with the rate of teenage births ($r = .340$), but is only weakly correlated with the mean income ($r = .190$) and with the rate of motor vehicle deaths ($r = .245$).

The task for multiple regression is to see how the independent variables predict the dependent variable *together*. If the correlation between mean per capita income and the rate of births to teenagers is $-.700$, that means that the independent variable accounts for 49% ($-.700^2$) of the variance in the dependent variable. And if the correlation between the rate of violent crimes and the rate of births to teenagers is $.340$, then the independent variable accounts for 11.56% ($.340^2$) of the variance in the dependent variable.

We can't just add these variances-accounted-for together, though, because the two independent variables are related to each other—each of the independent variables accounts for some variance in the other.

Table 21.17 shows what the output looks like from SYSTAT® when I asked the program to calculate the multiple correlation, called R, for INCOME and VIOLRATE with TEENBIRTH as the dependent variable.

Table 21.17 tells us that the regression equation is:

$$\text{TEENBIRTH} = 28.096 + (.006 \times \text{VIOLRATE}) + (-.001 \times \text{INCOME})$$

For example, the violence rate for Wisconsin was 281 crimes per 100,000 residents in 1995 and the average income in Wisconsin was \$21,184 in 1996.

TABLE 21.17
Multiple Regression Output from SYSTAT®

Dep Var: TEENBIRTH N: 50 Multiple-R: 0.850 Squared multiple-R: 0.722						
Adjusted squared multiple-R: 0.710 Standard error of estimate: 1.829						
<i>Effect</i>	<i>Coefficient</i>	<i>Std Error</i>	<i>Std Coef</i>	<i>Tolerance</i>	<i>t</i>	<i>P (2 Tail)</i>
CONSTANT	28.096	1.822	0.0	.	15.423	0.000
VIOLRATE	0.006	0.001	0.491	0.964	6.269	0.000
INCOME	-0.001	0.000	-0.793	0.964	-10.130	0.000

The regression equation predicts that the teenage birth rate for Wisconsin will be 8.6. The actual rate was 10.6 in 1996. The mean rate of teenage births for the 50 U.S. states was 12.9. The multiple regression, then, makes a better prediction than the mean: The difference between 10.6 and 12.9 is 2.3, while the difference between 8.6 and 10.6 is 2.0.

If you work out all the differences between the predictions from the *multiple* regression equation and the predictions from the *simple* regression equation involving *just* the effect of income on teenage births, the difference in the predictions will be the difference between accounting for 49% of the variance vs. accounting for 72.2% of the variance in the dependent variable.

Details of the Regression Equation

In this example, y is TEENBIRTH, x_1 is INCOME and x_2 is VIOLRATE. $R^2_{1\cdot23}$ is the variance in TEENBIRTH (the 1 in the subscript) accounted for by *both* INCOME and VIOLRATE (the 2 and 3 in the subscript). This relation is:

$$R^2_{1\cdot23} = r^2_{12} + r^2_{1(3\cdot2)}$$

which we can read as follows: The total variance in TEENBIRTH equals the variance accounted for by INCOME, plus the variance accounted for by VIOLRATE once the effect of INCOME has been accounted for. Calculating $r^2_{1\cdot23}$, then:

$$R^2_{1\cdot23} = r_{13} - (r_{12})(r_{32}) = .340 - (-.700)(.190) = .473$$

Taking the partial contribution of this relation to the dependent variable:

$$\frac{.473}{\sqrt{1 - .190^2}} = .481776$$

The contribution of this correlation to the variance of TEENBIRTH is $.481776^2$, or $.23211$. INCOME accounts for 49% of the variance in TEENBIRTH. Adding contributions, we get $.49 + .23211$, or 72.2%, which is the squared multiple-R in table 21.17.

There are three coefficients in the regression equation: a , b_1 , and b_2 . These are the coefficients in formula 21.3. Each of the b coefficients is the product of the standardized regression coefficient for each independent variable with the ratio of the standard deviation of the independent variable to the standard deviation of the dependent variable.

The standardized coefficient for the relation between (TEENBIRTH) and (INCOME) is:

$$b(x_1, x_2) = \frac{r_{12} - (r_{13})(r_{23})}{1 - (r_{23})^2}$$

for INCOME (x_2), $b = -.793$
 for VIOLRATE (x_3), $b = -.491$

These figures are given in table 21.17 as the “Std. Coef.,” or **standardized coefficients**.

The <i>SD</i> for the mean of INCOME is:	3,074,969
The <i>SD</i> for the mean of TEENBIRTH is:	3.399
The <i>SD</i> for the mean of VIOLRATE is:	269.225

Thus:

$$b_1 = (-.793)(3,074.969)/3.399 = -.001$$

and:

$$b_2 = (.491)(269.225)/3.399 = .006$$

These figures, and b_1 and b_2 , are given in table 21.17 as the “Coefficients,” and are the **unstandardized regression coefficients**. The method for calculating the value for a in the multiple regression equation is beyond the scope of this book. (For more about deriving multiple regression equations, consult Pedhazur [1997] or Gujarati [1995].)

But there’s more. If we add up the variances accounted for by the zero-order correlations of INCOME and VIOLRATE on TEENBIRTH, we get $49\% + 11.56\% = 60.56\%$. According to the results in table 21.17, however, the income in a state and the rate of violent crimes *together* account for 72.2% of the variance in teenage births. In other words, the two variables acting together account for *more* than they do separately, and this is the case despite the fact that the independent variables are moderately correlated ($r = .340$) with each other.

In fact, INCOME explains 43.82% of the variance in motor vehicle deaths ($r = -.662$ and r^2 is .4382) and VIOLRATE explains 6% of the variance in motor vehicle deaths ($r = .245$ and r^2 is .0600). *Together*, though, INCOME and VIOLRATE have a multiple-R of .762 and an R^2 of .581. Here again, the two variables explain more variance working together than they explain working separately.

In other words, it's the complex association of per capita income *and* the level of violence that explains so much variance in *both* the rate of teenage births and the rate of motor vehicle deaths. It turns out that lots of things are best explained by a series of variables acting together.

Caution: Automated Regression Now Available Everywhere

All of the major programs for statistical analysis today produce what is called a **stepwise multiple regression**. You specify a dependent variable and a series of independent variables that you suspect play some part in determining the scores of the dependent variable.

The program looks for the independent variable that correlates best with the dependent variable and then adds in the variables one at a time, accounting for more and more variance, until all the specified variables are analyzed, or until variables fail to enter because incremental explained variance is lower than a preset value, say, 1%. (There are even programs that test all possible path analysis equations . . . more on path analysis in a bit.)

Stepwise multiple regression is another one of those controversial hot topics in data analysis. Some people feel strongly that it is mindless and keeps you from making your own decisions about what causes what in a complex set of variables. It's rather like the significance test controversy and the shotgun controversy I discussed in chapter 20.

My take on it is the same: Learn to use all the tools and make your own decisions about the meaning of your findings. It's *your* responsibility to do the data processing and it's *your* responsibility to determine the meaning of your findings. Don't let robots take over any of your responsibilities. And don't be afraid to use all the hot new tools, either.

Examples of Multiple Regression

Graves and Lave (1972) modeled the starting wage of Navajo men who moved from the reservation in the 1960s to Denver. Graves and Lave started with 19 variables that they thought might relate in some way to starting wage—things like age, marital status, father's occupation, proficiency in

English, prior urban experience, internal vs. external control (feeling that you do or don't control what happens to you), and so on. Using stepwise multiple regression, Graves and Lave eliminated variables (like English proficiency, which vanished when they controlled for education) and settled on the following equation:

$$\text{STARTING WAGE} = .87 + 10 \text{ EDUC} + .08 \text{ VT} + \\ .13 \text{ PW} + .17 \text{ MARSTAT} + .11 \text{ FO}$$

where EDUC is the number of years of education beyond 10 (education only made a difference in starting wage after the 10th year), VT is the level of vocational training (skilled, like carpentry, vs. semiskilled, like house painting), PW is the highest premigration wage a man earned on the reservation, MARSTAT is marital status (married or unmarried), and FO is father's occupation (whether the migrant's father was a wage laborer or not—many Navajos in those days were sheep herders).

The equation says: Start with 87 cents an hour (remember, this was the mid-1960s); add 10 cents an hour for every year of education beyond 10; add 8 cents if the man had strong vocational skills; add 13 cents for each dollar the man had earned in his best job before migrating (in those days, the median premigration wage was \$1.50 an hour), add 17 cents if the man is married, and add 11 cents if his father had worked for wages.

R^2 for this equation is .54, which means that the equation accounts for 54% of the variance in starting wage.

Now that you know how it works, here are two more examples of multiple regression. John Poggie (1979) was interested in whether the beliefs of Puerto Rican fishermen about the causes of success in fishing were related to their actual success in fishing. He measured success by asking six key informants to rank 50 fishermen on this variable. Since his research was exploratory, he had a wide range of independent variables, three of which he guessed were related to fishing success: the fishermen's expressed orientation toward delaying gratification (measured with a standard scale), their boat size, and their years of experience at the trade.

The deferred gratification measure accounted for 15% of the variance in the dependent variable; years of experience accounted for another 10%; and boat size accounted for 8%. Together, these variables accounted for 33% of the variance in the success variable.

Emmanuel Mwangi (1986) studied the decision by farmers in Malawi to devote part of their land to growing new cash crops (tobacco and hybrid maize) rather than planting only the traditional crop, called "maize of the ancestors." His units of analysis were individual farms; his dependent variable was the ratio of land planted in tobacco and hybrid maize to the total land

under plow. And his eight independent variables were: the total cultivated land area, in hectares; the number of years a farmer was experienced in using fertilizers; whether the farming household usually brewed maize beer for sale; whether farmers owned any cattle at all; whether farmers had had any training in animal husbandry practices from the local extension agents; whether the family had an improved house (this required an index consisting of items such as a tin roof, cement floor, glass windows, and so on); whether the farmer owned a bicycle; and whether the farmer owned a plow and oxcart. These independent variables together accounted for 48% of the variance in the dependent variable.

On Explaining Just a Little of Something

In social science research, multiple regression (including path analysis, which is coming up next) typically accounts for between 30% and 50% of the variance in any dependent variable, using between two and eight independent variables. Like the equation from Graves and Lave (1972) above, you'll see plenty of regression equations with individual variables that explain 10% or less of what people are trying to understand. Does accounting for 10% of the variance in what you're interested in seem feeble? Consider:

1. In 2005, the average white male baby had a life expectancy at birth of 75.4 years in the United States, or 27,540 days. The life expectancy at birth for the average African American male baby was 69.9 years, or 25,532 days. The *difference* is 2,008 days. (In 1970, the figures were 67.1 years vs. 60.0 years, or a difference of 2,593 days.)
2. There were approximately 2.4 million births in Mexico in 2003, and around 52,000 infant deaths—that is, about 22.5 infant deaths per 1,000 live births. Compare these figures to the United States, where there were 4.1 million births and approximately 28,000 infant deaths, or about 6.8 per 1,000 live births. If the infant mortality rate in Mexico were the same as that in the United States, the number of infant deaths would be about 16,000 instead of 52,000. The *difference* would be 37,000 infant deaths.

Suppose you could account for just 10% of the *difference* in longevity among white and African American men in the United States (201 days) or 10% of the *difference* between the United States and Mexico in infant deaths (3,700 babies). Would that be worth doing? In my view, the most important contribution a social scientist can make to ameliorating a social problem is to be right about what causes it, and “accounting for variance” is part of causal modeling.

Path Analysis

Path analysis is a particular application of multiple regression. In multiple regression, we know (1) which independent variables help to predict some dependent variable and (2) how much variance in the dependent variable is explained by each independent variable. But multiple regression is an inductive technique: It does not tell us which are the antecedent variables, which are the intervening variables, and so on.

Path analysis is the application of multiple regression for testing conceptual models of multivariate relations—that is, for testing specific theories about how the independent variables in a multiple regression equation may be influencing each other—and how this ultimately leads to the dependent variable outcome.

The method was developed by the geneticist Sewall Wright in 1921 and became very popular in the social sciences in the 1960s (see Duncan 1966). It fell out of favor for a while (isn't it nice to know that even in statistics there are fads and fashions?), but it's making a comeback as full-featured statistics packages become more widely available.

I rather like the method because it depends crucially on the researcher's best guess about how a system of variables really works. It is, in other words, a nice combination of quantitative and qualitative methods. Here's an example.

John Thomas (1981) studied leadership in Niwan Witz, a Mayan village. He was interested in understanding what causes some people to emerge as leaders, while others remain followers. From existing theory, Thomas thought that there should be a relation among leadership, material wealth, and social resources. He measured these complex variables for all the household heads in Niwan Witz (using well-established methods) and tested his hypothesis using Pearson's *r*. Pearson correlations showed that, indeed, in Niwan Witz leadership is strongly and positively related to material wealth and control of social resources.

Since the initial hypothesis was supported, Thomas used multiple regression to look at the relation of leadership to *both* types of resources. He found that 56% of the variance in leadership was explained by just three variables in his survey: wealth (accounting for 46%), family size (accounting for 6%), and number of close friends (accounting for 4%). But, since multiple regression does not, as Thomas said, "specify the causal structure among the independent variables" (ibid.:132), he turned to path analysis.

From prior literature, Thomas conceptualized the relation among these three variables as shown in figure 21.2. He felt that leadership, *L*, was caused by all three of the independent variables he had tested, that family size (*fs*) influ-

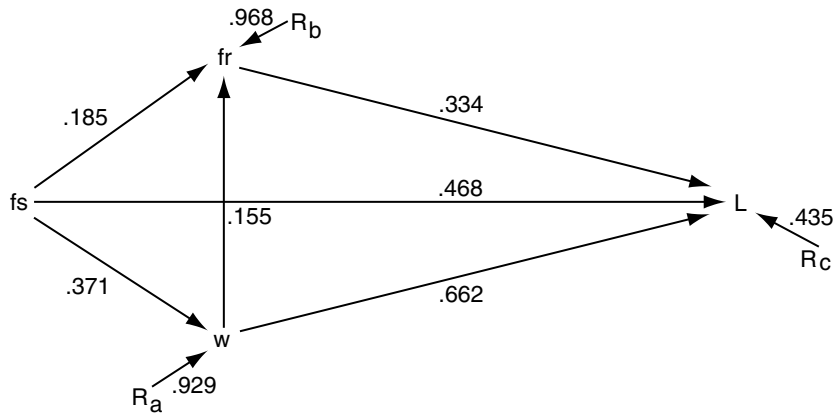


Figure 21.2. Path analysis of effects of wealth, friendship, and family size on leadership in Niwan Witz.

SOURCE: J. S. Thomas, “The Economic Determinants on Leadership on a Tojalabal Maya Community,” *American Ethnologist* Vol. 8, p. 1, 1981. Reproduced by permission of the American Anthropological Association. Not for further reproduction.

enced both wealth (*w*) and the size of one’s friendship network (*fr*), and that wealth was a factor in determining the number of one’s friends.

The **path coefficients** in figure 21.2 are **standardized values**: They show the influence of the independent variables on the dependent variables in terms of standard deviations. The path coefficients in figure 21.2, then, show that “a one standard deviation increase in wealth produces a .662 standard deviation increase in leadership; a one standard deviation increase in family size results in a .468 standard deviation increase in leadership; and so on” (Thomas 1981:133). (For details about how path coefficients are determined, consult a textbook in multivariate analysis, like Kelloway 1998.)

Four things are clear from figure 21.2: (1) Among the variables tested, wealth is the most important cause of leadership in individuals. (2) Family size has a moderate causal influence on wealth (making wealth a dependent, as well as an independent variable in this system). (3) The size of a person’s friendship network is only weakly related to either family size or wealth. (4) The combined direct and indirect effects of family size, wealth, and friendship network on leadership account for 56.5% ($1 - .435$) of the variance in leadership scores for the household heads of Niwan Witz.

Thomas concludes from this descriptive analysis that if you want to become a leader in the Mayan village of Niwan Witz, you need wealth, and the best way to get that is to start by having a large family.

Path analysis lets you test a particular theory about the relations among a system of variables, but it doesn't produce the theory; that's *your* job. In the case of Niwan Witz, for example, Thomas specified that he wanted his path analysis to test a particular model in which wealth causes leadership. The results were strong, leading Thomas to reject the null hypothesis that there really is no causal relation between wealth and leadership. But even the strong results that Thomas got don't prove anything. In fact, Thomas noted that an alternative theory is plausible. It might be that leadership in individuals (wherever they get it from) causes them to get wealthy rather than the other way around.

Path analysis often serves as reality therapy for social scientists. It's fun to build conceptual models—to think through a problem and hypothesize how variables are linked to each other—but our models are often much more complicated than they need to be. Path analysis can help us out of this fix. We can test several plausible theories and see which is most powerful. But in the end, you're left out there, all by yourself, defending your theory on the basis of whatever data are available right now.

Multicollinearity

Multivariate models are subject to a problem that simply can't exist when you have one independent variable: Independent variables can be correlated. In fact, when two variables both strongly predict a third, you'd expect the first two to be correlated. This **multicollinearity** means that you may not be able to tell the influence of one independent variable *free from the influence of the independent variables with which it is correlated*.

One way to avoid this problem is to conduct true experiments and nothing but true experiments. By assigning research participants randomly to control and experimental groups, we ensure that any correlation between independent variables is the result of chance. Nice work if you can get it, but most of what we want to study in anthropology requires fieldwork or survey research where things get messy.

Fortunately, multicollinearity distorts the findings of multiple regression only occasionally, and multicollinearity problems show up very clearly. The most dramatic sign of a multicollinearity problem is when your statistical program tells you that it cannot solve the equation. Some statistical programs calculate a **condition index** that diagnoses multicollinearity. A condition index between 10 and 30 signals moderate to strong multicollinearity, and an index over 30 signals a severe multicollinearity problem (Gujarati 1995).

But basic regression output shows the most common sign of a multicollin-

earity problem: Low probabilities for zero-order correlations between a series of independent variables and one dependent variable remain low in multivariate models when the variables are entered separately, but the probabilities for all the variables rise when you put them in the model together.

For example, an old puzzle, going all the way back to Durkheim (1951 [1897]), that remains of continuing interest is how to account for different rates of suicide across political units (countries, provinces, counties, cities). Table 21.18 shows the zero-order correlations among four variables for the 50

TABLE 21.18
Zero-Order Correlations of Four Variables That Appear to Be Related

	<i>SUICIDE</i>	<i>BTU</i>	<i>INCOME</i>	<i>OVER65</i>
<i>SUICIDE</i>	1.00			
<i>BTU</i>	0.437	1.00		
<i>INCOME</i>	-0.520	-0.330	1.00	
<i>OVER 65</i>	-0.367	-0.454	0.093	1.00

SOURCE: *SUICIDE* in 1994, Table 132, *Statistical Abstract of the United States* (1995). *BTU* in 1994, Table 925, *Statistical Abstract of the United States* (1997). *INCOME* in 1995, Table 698, *Statistical Abstract of the United States*. *OVER 65* in 1995, Table 34, *Statistical Abstract of the United States* (1996).

states of the United States plus Washington, D.C. *SUICIDE* is the number of deaths by suicide, per 100,000 in 1994; *BTU* is the average energy consumption, per person, in millions of BTUs in 1995; *INCOME* is the average personal income in 1995; and *OVER65* is the percentage of people who were over 65 in 1996.

Notice that the rate of suicide (the ninth-leading cause of death in the United States) is correlated with each of the three independent variables. The higher the level of energy consumption (i.e., the worse the weather), the higher the rate of suicide. The higher the average personal income, the lower the rate of suicide. And the higher the proportion of old people, the *lower* the rate of suicide. (Suicide is the third-leading cause of death, after accidents and homicide, among Americans 15–24 years of age, but is hardly a factor among the elderly.)

Furthermore, each of these zero-order correlations is statistically very significant. Table 21.19 shows the regression between each of these independent variables separately and the rate of suicide. The *p*-value in each case is very low, indicating a highly significant statistical relation.

Table 21.20 shows what happens when we use multiple regression to check for spurious relations. If we put *INCOME* and *BTU* into the regression together, both variables remain statistically significant. If we put *INCOME* and *OVER65* into the regression, both variables remain significant. Thus, both

TABLE 21.19
Separate Regressions of Three Independent Variables with SUICIDE

Dep Var: SUICIDE N: 51 Multiple R: 0.520 Squared multiple R: 0.270
Adjusted squared multiple R: 0.255 Standard error of estimate: 3.112

<i>Effect</i>	<i>Coefficient</i>	<i>Std Error</i>	<i>Std Coef</i>	<i>Tolerance</i>	<i>t</i>	<i>P (2 Tail)</i>
CONSTANT	24.789	2.799	0.000	.	8.856	0.000
INCOME	-0.001	0.000	-0.520	1.000	-4.257	0.000

Dep Var: SUICIDE N: 51 Multiple R: 0.437 Squared multiple R: 0.191
Adjusted squared multiple R: 0.174 Standard error of estimate: 3.277

<i>Effect</i>	<i>Coefficient</i>	<i>Std Error</i>	<i>Std Coef</i>	<i>Tolerance</i>	<i>t</i>	<i>P (2 Tail)</i>
CONSTANT	9.309	1.184	0.000	.	7.860	0.000
BTU	0.010	0.003	0.437	1.000	3.397	0.001

Dep Var: SUICIDE N: 51 Multiple R: 0.372 Squared multiple R: 0.139
Adjusted squared multiple R: 0.121 Standard error of estimate: 3.381

<i>Effect</i>	<i>Coefficient</i>	<i>Std Error</i>	<i>Std Coef</i>	<i>Tolerance</i>	<i>t</i>	<i>P (2 Tail)</i>
CONSTANT	21.509	3.061	0.000	.	7.026	0.000
OVER65	-0.667	0.238	-0.372	1.000	-2.807	0.007

TABLE 21.20
Regressions of the Three Pairs of Independent Variables with SUICIDE

Dep Var: SUICIDE N: 51 Multiple R: 0.591 Squared multiple R: 0.349
Adjusted squared multiple R: 0.322 Standard error of estimate: 2.969

<i>Effect</i>	<i>Coefficient</i>	<i>Std Error</i>	<i>Std Coef</i>	<i>Tolerance</i>	<i>t</i>	<i>P (2 Tail)</i>
CONSTANT	20.038	3.318	0.000	.	6.038	0.000
INCOME	-0.000	0.000	-0.421	0.891	-3.417	0.001
BTU	0.007	0.003	0.298	0.891	2.413	0.020

Dep Var: SUICIDE N: 51 Multiple R: 0.610 Squared multiple R: 0.372
Adjusted squared multiple R: 0.346 Standard error of estimate: 2.915

<i>Effect</i>	<i>Coefficient</i>	<i>Std Error</i>	<i>Std Coef</i>	<i>Tolerance</i>	<i>t</i>	<i>P (2 Tail)</i>
CONSTANT	31.282	3.502	0.000	.	8.933	0.000
INCOME	-0.001	0.000	-0.490	0.991	-4.264	0.000
OVER65	-0.564	0.202	-0.321	0.991	-2.798	0.007

BTU and OVER65 show effects on suicide rates that are independent of the effects of INCOME.

But look what happens when we enter all three independent variables into the regression. Table 21.21 shows the result: INCOME *remains* statistically significant, but *neither* BTU *nor* OVER65 are.

TABLE 21.21
Regression of All Three Independent Variables and SUICIDE

<i>Effect</i>	<i>Coefficient</i>	<i>Std Error</i>	<i>Std Coef</i>	<i>Tolerance</i>	<i>t</i>	<i>P (2 Tail)</i>
CONSTANT	26.898	4.806	0.000	.	5.597	0.000
INCOME	-0.000	0.000	-0.434	0.889	-3.613	0.001
BTU	0.004	0.003	0.183	0.714	1.362	0.180
OVER65	-0.439	0.228	-0.245	0.793	-1.927	0.060

If BTU remained statistically significant while OVER65 did not, we could conclude that OVER65 had no effect on suicide independently of INCOME and BTU. Likewise, if OVER65 remained statistically significant while BTU did not, we could conclude that BTU had no effect on suicide independently of INCOME and OVER65. In this instance, however, we can only conclude that we have a multicollinearity problem that obscures our findings. We can't tell if either BTU or OVER65 has effects on suicide independently of INCOME. (In fact, the condition index for this regression is 27, which indicates strong to severe multicollinearity.)

Factor Analysis

Factor analysis is based on the simple and compelling idea that if things we observe are correlated with each other, they must have some underlying variable in common. Factor analysis refers to a set of techniques for identifying and interpreting those underlying variables.

For example, people in the United States who are in favor of gun control are likely (but not guaranteed) to be in favor of: (1) a woman's right to an abortion; (2) participation by the U.S. military in overseas peacekeeping missions of the United Nations; and (3) affirmative action in college admissions. People who are against gun control are likely to favor: (1) restrictions on abortion; (2) less involvement of the United States in UN peacekeeping missions; and (3) curbs on affirmative action.

Of course, people are free to mix and match their opinions on any of these issues, but overall, there is a strong association of opinions across these issues. Factor analysis assumes that this association is the result of an underlying, hypothetical variable, a general attitude orientation. We can use a set of statistical techniques to identify both the existence and the content of this underlying variable. Naming the variable is a strictly qualitative exercise.

Many studies have demonstrated the existence of an underlying value orientation usually called “liberal vs. conservative.” Liberals and conservatives exhibit packages of attitudes about things like personal freedom (as with the right to own a handgun or the right to an abortion), about foreign policy (as with involvement in peacekeeping missions), and about domestic policies (as with affirmative action).

This idea of multidimensional supervariables that underlie a set of observations was first articulated in 1904 by Charles E. Spearman (he for whom Spearman’s rank-order correlation coefficient is named—see chapter 20). Spearman noticed that the scores of students on various exams (classics, French, English, math, etc.) were correlated. He suggested that the exam scores were correlated with each other because they were all correlated with an underlying factor, which he labeled *g*, for general intelligence.

The single-factor theory of intelligence has been repudiated many times since then, as have been the IQ tests that were developed to measure intelligence as a single factor. But the idea of factors—multidimensional variables—that underlie and give rise to a set of correlated events became one of the most important developments in all the social sciences.

The data for factor analysis are often attitude items, like the kind you see in Likert scales, but can just as well be about artifacts (movies, songs, buildings, cars, brands of beer . . .), people (movie stars, politicians, fashion models, criminals, classical musicians . . .), or even countries. Factor analysis is used across the social sciences in data reduction—to explore large data sets with dozens or even hundreds of variables in order to extract a few variables that tell a big story.

In political science, for example, Lieske (1993) partitioned the 3,164 counties in the United States into 10 distinctive regional subcultures by looking for common, underlying factors in the correlations among 45 racial, ethnic, religious, and social structural variables. In social psychology, Stein et al. (1991) used factor analysis to test the influence of ethnicity, socioeconomic status, and various anxieties as barriers to the use of mammography among white, African American, and Hispanic women. Factor analysis confirmed the influence of five kinds of anxiety: fear of radiation, fear of pain, embarrassment about breast exams, anxiety about what might be found, and concerns about cost.

Factor Analysis and Scales

As I mentioned in chapter 12, factor analysis is widely used across the social sciences in building reliable, compact scales for measuring social and psychological variables. Suppose, for example, you are interested in attitudes toward gender role changes among women. You suspect that the underlying forces of role changes are related to premarital sexuality, working outside the home, and the development of an independent social and economic life among women. You make up 50 attitudinal items and collect data on those items from a sample of respondents.

Factor analysis will help you decide whether the 50 items you made up really test for the underlying forces you think are at work. If they do, then you could use a few benchmark items—the ones that “load high” on the factors—and this would save you from having to ask every respondent about all 50 items you made up. You would still get the information you need—or much of it, anyway. How much? The amount would depend on how much variance in the correlation matrix each of your factors accounted for.

The notion of variance is very important here. Factors account for chunks of variance—the amount of dispersion or correlation in a correlation matrix. Factors are extracted from a correlation matrix in the order of the amount of variance that they explain in the matrix. Some factors explain a lot of variance, while others may be very weak and are discarded by researchers as not being useful. In a dense matrix, only one or a few factors may be needed to account for a lot of variance, while in a dispersed matrix, many factors may be needed.

The most common statistical solution for identifying the underlying factors in a correlation matrix is called the **orthogonal solution**. In orthogonal factor analyses, factors are found that have as little correlation with each other as possible. Other solutions that result in intercorrelated factors are also possible (the various solutions are options that you can select in all the major statistical packages, like SAS®, SYSTAT®, and SPSS®). Some researchers say that these solutions, although messier than orthogonal solutions, are more like real life.

So-called **factor loadings** are the correlations between the factors and the variables that are subsumed by, or appear to be components of, factors. All the old variables “load” on each new factor. The idea is to establish some cutoff below which you would not feel comfortable accepting that an old variable loaded onto a factor. Many researchers use 0.50 as the cutoff, and look at loadings of 0.30–.49 as worth considering. Some researchers insist that a variable should load at least 0.60 before accepting it as an unambiguous component of a factor and look at variables that load between 0.30 and 0.59 as worth considering.

Once you have a list of variables that load high on a factor (irrespective of

sign, plus or minus), you look at the list and decide what the factor *means*. An example should make all this a lot clearer.

Handwerker's Domestic Activities Scale

Penn Handwerker (1996a, 1998) used factor analysis to test whether the construct of “domestic cooperation” on Barbados was unidimensional. He asked a random sample of 428 Barbadian women whether their husband or boyfriend helped with any of the following: cooking, washing clothes, washing dishes, bathing children, taking children places, and caring for children. To put these items of domestic cooperation in context, he also asked each woman whether her husband or boyfriend was *expected* to treat her as an equal and whether her husband or boyfriend *did, in fact*, treat her as an equal.

Table 21.22 is a schematic of Handwerker's data matrix. The eight variables are labeled COOK, WASH, DISH, BATHE, TAKE, CARE, EQUAL1, and EQUAL2.

TABLE 21.22
Schematic of Handwerker's Profile Matrix

<i>ID</i>	<i>COOK</i>	<i>WASH</i>	<i>DISH</i>	<i>BATH</i>	<i>TAKE</i>	<i>CARE</i>	<i>EQUAL1</i>	<i>EQUAL2</i>
1	1	0	1	1	1	0	1	0
2	0	1	0	0	1	0	1	1
3	1	0	0	0	1	1	1	0
.								
.								
438	1	1	0	1	1	0	1	1

SOURCE: W. P. Handwerker, adapted from “Constructing Likert Scales: Testing the Validity and Reliability of Single Measures of Multidimensional Variables,” *Cultural Anthropology Methods Journal*, Vol. 8, no. 1, p. 2, 1996a. Reprinted by Permission of Sage Publications.

Table 21.22 is a profile matrix (see figure 16.1a), but factor analysis is done on a similarity matrix (see figure 16.1b). If Handwerker had asked the women: “On a scale of 1–5, how much does your husband or boyfriend help you with the cooking,” the entries in table 21.22 would have been 1–5. With that kind of data, a factor analysis program would turn the profile matrix into a similarity matrix by calculating Pearson's *r* for all possible pairs of columns. But Handwerker asked the women Yes/No questions (that's why table 21.22 contains only 0s and 1s).

One way to turn a 1/0 profile matrix into a similarity matrix is to calculate the percentage of matches for all possible pairs of columns. That is, when two

columns have a 1 or a 0 in the same row, count that as a hit. When two columns have different entries in the same row, count that as a miss. Then, count up all the hits and divide by 428 (the number of possible hits for 428 respondents).

This results in what's called a **simple matching coefficient**. (There are other kinds of matching coefficients, but I won't go into them here.) The result is an 8×8 similarity matrix. Table 21.23 shows the results of Handwerker's factor analysis of that matrix.

TABLE 21.23
Factor Loadings for Handwerker's Data

<i>Variable</i>	<i>Factor 1 Domestic</i>	<i>Factor 2 Children</i>	<i>Factor 3 Affection</i>
COOK	0.893	0.232	0.161
WASH	0.895	0.121	0.058
DISH	0.824	0.329	0.191
BATHE	0.795	0.324	0.194
TAKE	0.291	0.929	0.159
CARE	0.307	0.922	0.175
EQUAL1	0.188	0.203	0.734
EQUAL2	0.091	0.066	0.854
Explained Variance	39.15	25.516	17.841

SOURCE: P. Handwerker, "Constructing Likert Scales: Testing the Validity and Reliability of Single Measures of Multidimensional Variables," *Cultural Anthropology Methods Journal*, Vol. 8, no. 1, p. 2, 1996a. Reprinted with Permission of Sage Publications.

In reading the output from factor analysis, we look for items that load high on each factor. For me, that means loadings of at least .60. From table 21.23, it's clear that the domain of "domestic cooperation" is *not* unidimensional. In fact, it has three dimensions. Interpreting the results in table 21.23, it seemed to Handwerker that Factor 1 had something to do with "household chores." He labeled that factor "Domestic." Factor 2, he thought, comprised "chores associated with children," so he labeled it "Children." In open-ended interviews, Barbadian women interpreted the third factor as being about affection and empowerment within families, so Handwerker labeled it "Affection."

Over the next few years, Handwerker had the opportunity to refine and test his scale on two more Caribbean islands, Antigua and St. Lucia. He dropped EQUAL1 because it was redundant with EQUAL2 and he dropped the question about "caring for children in other ways" (CARE) because respondents told him that it was ambiguous.

Handwerker also added four new questions—things that had come up in

open-ended interviews as important to women: (1) Does your partner take responsibility for the children for an evening or an afternoon when you have something to do? (TIME) (2) Does your partner take time off from work to share responsibility for children who are sick? (SICK) (3) Does your partner talk with you and respect your opinion? (TALK) (4) Does your partner spend his free time with you? (FREET)

Table 21.24 shows the results for Antigua and St. Lucia.

TABLE 21.24
Handwerker's Data from Antigua and St. Lucia

Variable	Antigua			St. Lucia		
	Domestic	Children	Affection	Domestic	Children	Affection
COOK	0.793	0.317	0.106	0.783	0.182	0.161
WASH	0.845	0.163	0.090	0.818	0.056	0.018
DISH	0.791	0.320	0.203	0.781	0.197	0.173
BATHE	0.493	0.654	0.207	0.695	0.584	0.090
TAKE	0.271	0.738	0.289	0.246	0.790	0.196
TIME	0.253	0.820	0.258	0.228	0.833	0.061
SICK	0.210	0.786	0.151	0.121	0.742	-.067
FREET	0.100	0.299	0.802	-.010	0.691	0.319
EQUAL	0.147	0.155	0.898	0.177	0.117	0.883
TALK	0.142	0.200	0.883	0.113	0.112	0.909
Explained variance	26.460	26.459	25.051	24.188	27.905	18.191

SOURCE: W. P. Handwerker, "Constructing Likert Scales: Testing the Validity and Reliability of Single Measures of Multidimensional Variables," 1996a. *Cultural Anthropology Methods Journal*, Vol. 8, no. 1, p. 3, 1996a. Reprinted with Permission of Sage Publications.

There are at least four things to notice about table 21.24:

1. Despite the subtraction of some variables and the addition of others, the results from all three islands are very stable. COOK, WASH, and DISH are components of a single large factor across all three Caribbean countries.
2. While BATHE loads high on the domestic chore factor across all three islands, it also loads high on the children factor for two of the islands. This item should be dropped in the future because it does not reliably distinguish the two factors.
3. FREET loads on the children factor for St. Lucia, but it loads on the affection factor for Antigua. It turns out that Handwerker's assistant changed the wording for the FREET question slightly when she did the interviews on Antigua. Instead of asking: "Does your partner spend his free time with you?" she asked "Does your partner spend his free time with you or with your children?" (Handwerker 1996a:3). That little change apparently made enough of a difference in the responses to change the factor loading.

4. Across all three replications, the three factors in tables 21.23 and 21.24 account for 70%–80% of the variance in the original data matrix. That is, about three-fourths to four-fifths of the variance in the original data is accounted for by just three underlying variables (the three factors) rather than the full list of original variables. For women across the Caribbean, the construct of domestic cooperation is multidimensional and comprised of three subconstructs: sharing of everyday domestic chores, sharing of responsibilities for children, and affection from men as defined by being treated as an equal.

Multidimensional Scaling Analysis (MDS)

MDS is another multivariate data-reduction technique. Like factor analysis, it is used to tease out underlying relations among a set of observations. Also like factor analysis, MDS requires a matrix of measures of associations—e.g., a correlation matrix based on things like r , tau, gamma, etc. But unlike factor analysis, MDS does not require **metric data**.

When you measure something like how strongly people feel about something, the numbers you assign to their feelings don't have the same meaning as, say, numbers that express distances in kilometers or kilograms of game meat killed per month. These latter numbers are called “metric” because they are grounded in well-understood units of measurement.

Most attitude and cognition data are nonmetric. MDS is particularly useful for anthropologists, since a lot of the measurements we make are nonmetric. Also, MDS produces a graphic display of the relation among any set of items, whether those items are people, or things, or questions about attitudes. (For more about the theory behind MDS, see Romney et al. [1972].)

How MDS Works

Suppose you measure the association among three variables, A, B, and C, using Pearson's r . The association matrix for these three variables is in the *inside box* of table 21.25.

Clearly, variables A and C are more closely related to one another than are A and B, or B and C. You can represent this with a triangle, as in figure 21.3a.

In other words, we can place points A, B, and C on a plane in some position relative to each other. The distance between A and B is longer than that between A and C (reflecting the difference between .50 and .80); and the distance between B and C is longer than that between A and C (reflecting the difference between .40 and .80). The numbers in this graph are **similarities**: The lower the correlation, the longer the distance; the higher the correlation, the shorter the distance. (We won't consider negative correlations here.)

TABLE 21.25
Matrix of Association among Four Variables

	A	B	C	D
A	X	0.5	0.8	0.3
B		X	0.4	0.64
C			X	0.35
D				X

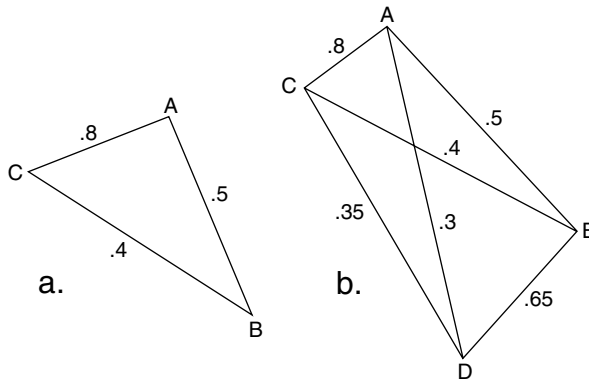


Figure 21.3. Two-dimensional plot of the relation among three variables (a) and among four variables (b).

With just three variables, it is easy to plot these distances in proper proportion to one another. For example, the physical distance between B and C is twice that of A and C in figure 21.3a. In fact, figure 21.3a contains *precisely* the same information as the inside box of table 21.25, but in graphic form.

With four variables, things get considerably more complicated. With four variables, there are *six* relations to cope with. These relations are shown in the large box of table 21.25. Only one two-dimensional graph (apart from rotations and enlargements) can represent the relative distances among the six relations in table 21.25. The graph is shown in figure 21.3b.

Figure 21.3b is a two-dimensional graph of six relations in *almost* proper proportions. When we have only two dimensions to work with, it is often impossible to achieve perfect proportionality in a graph of six relations. One way out of this is to depict the six relations in table 21.25 in three dimensions,

instead of in only two. The extra dimension would give us plenty of room to move around and we could better adjust the proportionality of the distances between the various pairs of variables.

In principle, you can perfectly represent the relations among n variables in $n - 1$ dimensions, so that any graph of six variables can be perfectly represented in five dimensions. But even a three-dimensional graph is hard to read. What would you do with a five-dimensional graph?

Most researchers specify a two-dimensional solution when they run an MDS computer analysis. MDS programs produce a statistic that measures the “stress” in the graph produced by the program. This is a measure of how far off the graph is from one that would be perfectly proportional. The lower the stress, the better the solution. This means that a cluster of variables in an MDS graph with low stress is likely to reflect some reality about the cognitive world of the people being studied.

A Physical World Example: Making Maps with MDS

A physical example, based on metric data, will make this clearer. Afterward, we’ll move on to a cognitive example, one based on nonmetric data. Table 21.26 shows the road distance in miles between all pairs of nine cities in the United States.

TABLE 21.26
Distances between Nine U.S. Cities (in Miles)

	<i>BOS</i>	<i>NY</i>	<i>DC</i>	<i>MIA</i>	<i>CHI</i>	<i>SEA</i>	<i>SF</i>	<i>LA</i>	<i>DEN</i>
Boston	0								
NY	206	0							
DC	429	233	0						
Miami	1504	1308	1075	0					
Chicago	963	802	671	1329	0				
Seattle	2976	2815	2684	3273	2013	0			
SF	3095	2934	2799	3053	2142	808	0		
LA	2979	2786	2631	2687	2054	1131	379	0	
Denver	1949	1771	1616	2037	996	1037	1235	1059	0

SOURCE: *Anthropac 4.0* and *Anthropac 4.0 Methods Guide*, by S. P. Borgatti, 1992a, 1992b. Reprinted with permission of the author.

Note two things about the numbers in this table. First, the numbers are **dis-similarities**, not similarities, because bigger numbers mean that things are farther apart—less like each other. Smaller numbers mean that things are close together—more similar. Similarity and dissimilarity matrices are known col-

lectively as **proximity matrices** because they tell you how close or far apart things are. The second thing to notice is that the numbers are reasonably accurate measures of a physical reality—distance between points on a map—so they are metric data.

In principle, there should be a two-dimensional solution with low stress that fits the numbers in this table. I used *Anthropac* (Borgatti 1992a) to run MDS on these data, and the solution is shown in figure 21.4.

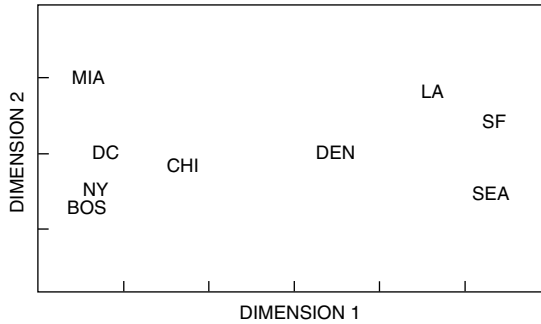


Figure 21.4. Two-dimensional MDS solution for the numbers in table 21.26.

Figure 21.4 looks suspiciously like a map of the United States. All nine cities are placed in proper juxtaposition to one another, but the map looks sort of upside-down and backward. If we could only flip the map over from left to right and from bottom to top. . . . Multidimensional scaling programs are notoriously unconcerned with details like this. So long as they get the juxtaposition right, they're finished. Figure 21.4 shows that the program got it right. (Obviously, you can rotate any MDS graph through some angle about any axis, and it will still have the same meaning. Think about a map of the surface of the Earth from the perspective of someone who is looking out from a hundred miles *inside* the Earth. It would be the same map that you're accustomed to, but flipped.)

This means that you can use MDS to create outline maps of your own. To map a village, you would measure the distance between all pairs of some set of points around the perimeter. A set of 10 points means 45 pairs and 45 measurements. Of course, the more pairs of points you measure, the more accurate the map will be, but, as you can see from figure 21.4, even a set of nine points (and 36 measurements) gets you a basic outline map of the United States.

A Cognitive World Example

We can use MDS to draw and explore **mental maps** of the items in a cultural domain. To show you how this works, table 21.27 shows the data from

TABLE 21.27
 Pile Sort Data for an Individual Proximity Matrix

<i>A List of 18 Fruits for a Pile Sort</i>	
1. Apple	10. Strawberry
2. Orange	11. Lemon
3. Papaya	12. Cantelope
4. Mango	13. Grapefruit
5. Peach	14. Plum
6. Blueberry	15. Banana
7. Watermelon	16. Avocado
8. Pineapple	17. Fig
9. Pear	18. Cherry
<i>One Person's Sorting of 18 Fruits</i>	
Pile #1: 2, 11, 13	
Pile #2: 1, 5, 9, 14, 17, 18	
Pile #3: 3, 4, 8, 15, 16	
Pile #4: 6, 10	
Pile #5: 7, 12	

one person's pile sort of 18 fruits (the names of the fruits are in the top half of the table and the contents of each pile the informant made are in the bottom half). Table 21.28 displays the pile-sort data as an **individual proximity matrix**.

This particular proximity matrix is a **similarity matrix**. (Recall that the matrix of cities was a dissimilarity matrix.) When the informant put items 2, 11, and 13 into a pile, he did so because he thought the items were similar. To indicate this, there is a 1 in the matrix where items 2 and 11 intersect; another 1 in the cell where items 2 and 13 intersect; and another 1 in the cell where 11 and 13 intersect. And similarly for Pile #2: there is a 1 in the 1–5 cell, the 1–9 cell, the 1–14 cell, and so on. There are zeros in all the cells that represent no similarity of a pair of items (for this informant) and zeros down the diagonal (we don't count items as being similar to themselves). Also, notice that if 11 is similar to 13, then 13 is similar to 11, so this particular matrix is also **symmetric**. The bottom and top halves of the matrix are identical.

The matrix in table 21.28 contains *exactly and only* the information in the bottom half of table 21.27. There are no numerical operations here. The individual proximity matrix in table 21.28 is just a convenient way to display the symmetric similarities implied by the pile sort data. Figure 21.5 is a multidimensional scaling of these similarity data, a visualization of how this infor-

TABLE 21.28
Similarity Matrix from One Person's Pile Sorting of the 18 Fruits

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
	AP	OR	PA	MA	PE	BL	WA	PI	PE	ST	LE	CA	GR	PL	BA	AV	FI	CH
1	APPLE	1	0	0	1	0	0	0	1	0	0	0	0	1	0	0	1	1
2	ORANGE	0	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0
3	PAPAYA	0	0	1	1	0	0	1	0	0	0	0	0	0	1	1	0	0
4	MANGO	0	0	1	1	0	0	1	0	0	0	0	0	0	1	1	0	0
5	PEACH	1	0	0	1	0	0	0	1	0	0	0	0	1	0	0	1	1
6	BLUEBERRY	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0
7	WATERMELON	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
8	PINEAPPLE	0	0	1	1	0	0	1	0	0	0	0	0	0	1	1	0	0
9	PEAR	1	0	0	0	1	0	0	1	0	0	0	0	1	0	0	1	1
10	STRAWBERRY	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0
11	LEMON	0	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0
12	CANTELOPE	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
13	GRAPEFRUIT	0	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0
14	PLUM	1	0	0	0	1	0	0	1	0	0	0	0	1	0	0	1	1
15	BANANA	0	0	1	1	0	0	1	0	0	0	0	0	0	1	1	0	0
16	AVOCADO	0	0	1	1	0	0	1	0	0	0	0	0	0	1	1	0	0
17	FIG	1	0	0	0	1	0	0	1	0	0	0	0	1	0	0	1	1
18	CHERRY	1	0	0	0	1	0	0	1	0	0	0	0	1	0	0	1	1

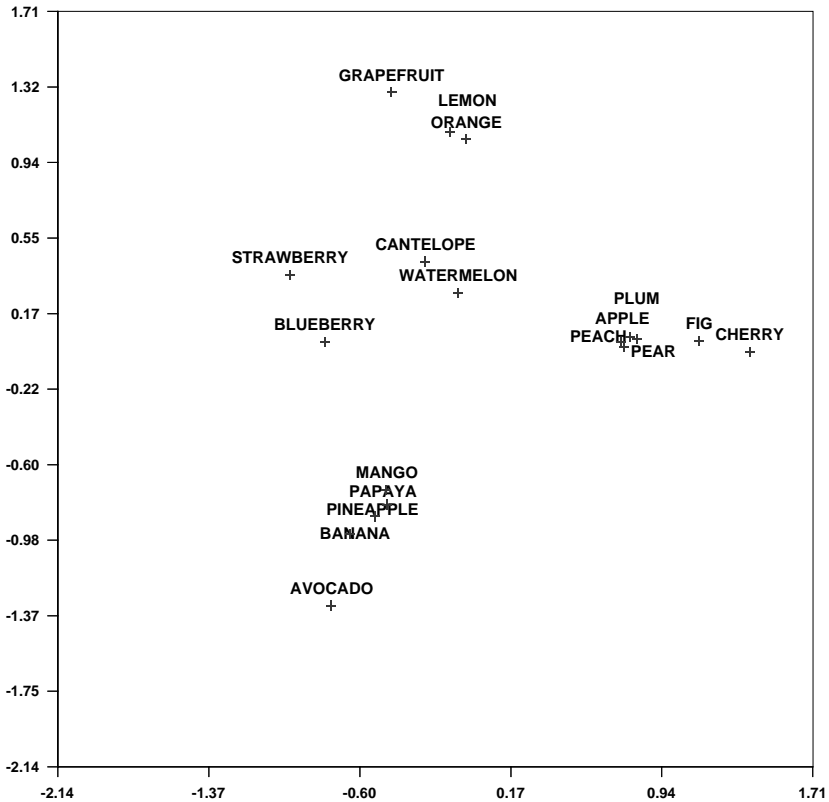


Figure 21.5. Two-dimensional MDS for 18 fruits sorted by one informant.

mant sees the similarities among these 18 fruits. It is, in other words, a **mental map**. Make no mistake here. The MDS graph is simply a *representation* of what the informant was thinking when he pile sorted those fruits. No one claims that MDS graphs are one-for-one maps of what's happening inside people's heads.

Now, suppose we collect a whole set of pile sort data from, say, 20 informants. That would produce 20 individual similarity matrices. If we stacked those matrices on top of one another we could look down through each cell, count up the number of 1s, and divide by 20. That would tell us the *percentage* of times that the 20 informants put a pair of items (fruits, for example) together in a pile. This would create an **aggregate proximity matrix**—a summary of percentages, rather than just hits, that tell us how similar, on average

each pair of items is for this set of informants. If we run the MDS on the aggregate matrix we'd get a mental map for the group.

Susan Weller (1983) asked 20 Guatemalan women to list as many illnesses as they could think of. Then she took the 27 most frequently named illnesses, put each named illness on a card, and asked 24 other women (some urban, some rural) to sort the cards into piles, according to similarity. Weller created an aggregate proximity matrix from the 24 individual matrices and did a multidimensional scaling to represent how her informants collectively perceived the 27 illnesses. Figure 21.6 shows the MDS graph for Weller's urban sample

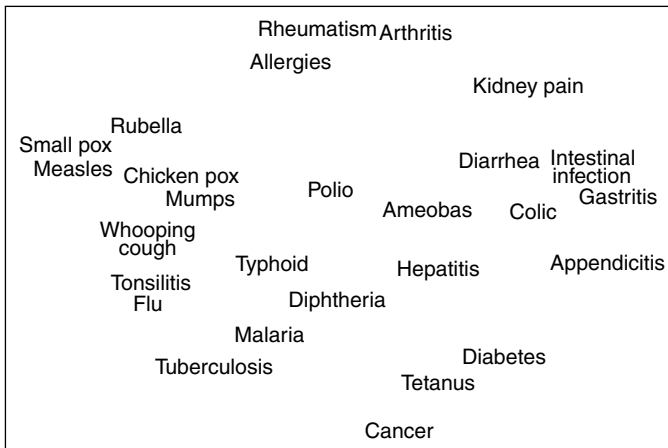


Figure 21.6. MDS representation of 27 illnesses for urban Guatemalan women.

SOURCE: S. Weller, "New Data on Intracultural Variability: The Hot-Cold Concept of Medicine and Illness," *Human Organization*, Vol. 42, pp. 249–257. © 1983. Reprinted with permission of the Society for Applied Anthropology.

of the pile-sort data. The MDS program converts similarities (percentages, in this case) into distances. The illness terms that were judged to be similar are closer together in figure 21.6, and the terms judged to be dissimilar by informants are farther apart.

This is the fun part. Look at figure 21.6. Keep looking at it. Think about the distribution of those items (the illness names). Do you see any pattern? We're looking for **arrays**, or **dimensions** of illnesses across the plot and **clumps**, or **clusters** of illnesses scattered around the plot. Since the MDS program was able to plot the items in two dimensions with acceptably low stress, we are looking for two arrays. There can be any number of clumps. Interpretation means figuring out what the dimensions and the clumps are.

Try naming the clusters in figure 21.6. There is a clump on the right that might be called "gastrointestinal disorders." On the left there is a clump of

“childhood disorders.” Those, at least, are the labels that struck Weller as appropriate. I agree with her intuition about this. What do *you* think?

How about the arrays, or dimensions? We’re accustomed to looking for vertical and horizontal axes in graphs, but this convention is irrelevant for interpreting MDS output. Dimensions in MDS graphs don’t necessarily have to be at right angles to each other. To me, it looks like informants distinguish between chronic and acute illnesses (from top to bottom in figure 21.6) and between infectious and noninfectious illnesses (from left to right). The fact that I can come up with labels, though, says absolutely nothing about whether I understand what is going on. It is possible to label anything, once you’re confronted with the task. This means that you must be especially careful in the use of factor analysis, MDS, and other dredging techniques that present you with something to explain.

On the other hand, the mere fact that I might make a mistake in my interpretation of the results doesn’t stop me from using these techniques. Interpretation of numerical results is always a brazen, flat-out qualitative exercise, a Rorschach test for social scientists—which is why I like it so much. Use every technique you can think of in data analysis, and let your experience guide your interpretation. Interpretation of results is where data analysis in all science ultimately becomes a humanistic activity. (For more examples of MDS in anthropology, see Johnson and Griffith 1995, Magaña et al. 1995, and Bolton and Vincke 1996.)

Cluster Analysis

Like factor analysis and MDS, cluster analysis is a descriptive tool for exploring relations among items in a matrix—for finding what goes with what. You start with a similarity matrix, like a matrix of Pearson correlation coefficients. If you factor the matrix, you find underlying variables that may encompass the variables in the original data. If you scale the matrix (MDS), you get a map that shows you graphically the relations among the items. Clustering tells you which items go together and in what order. And, of course, the output from all these analyses should agree, more or less. Thus, in the MDS example from Weller (1983) above, a cluster analysis would let us check the guess about that chunk of illnesses labeled “gastrointestinal disorders.”

To get an idea of where we’re going with this, take another look at figure 21.6. Does kidney pain go with arthritis-rheumatism-allergies group of chronic illnesses or does it go with the diarrhea etc. group of intestinal infections? Or is it somehow a cognitive bridge between these groups? This is the sort of thing we want cluster analysis to help us determine.

I'm going to explain cluster analysis in some detail. It's a very important descriptive tool, but as we go through the next few paragraphs, keep in mind two things: First, clustering is just a technique for finding the similarity chunks. It doesn't label those chunks. That part is also a Rorschach test: You stare at the output and decide what the meaning is. Second, as with so many methods, different treatments of your data produce different outcomes. The next few pages will make it very clear just how much *you and only you* are responsible for every choice you make in data analysis.

How Cluster Analysis Works

Consider the following example from de Ghett (1978:121):

1 3 7 9 14 20 21 25

This set of numbers has no meaning at all, so we can concentrate on the method of clustering them, without any interpretation getting in the way. When we get through with this example, we'll move on to a set of numbers that does have meaning. The distance between 1 and 3 is 2. The distance between 21 and 25 is 4. So, in a numerical sense, 1 and 3 are twice as similar to one another as 21 and 25 are to one another. Table 21.29 shows the **dissimilarity matrix** for these numbers.

TABLE 21.29
Dissimilarity Matrix for Clustering

	<i>1</i>	<i>3</i>	<i>7</i>	<i>9</i>	<i>14</i>	<i>20</i>	<i>21</i>	<i>25</i>
<i>1</i>	0							
<i>3</i>	2	0						
<i>7</i>	6	4	0					
<i>9</i>	8	6	2	0				
<i>14</i>	13	11	7	5	0			
<i>20</i>	19	17	13	11	6	0		
<i>21</i>	20	18	14	12	7	1	0	
<i>25</i>	24	22	18	16	11	5	4	0

SOURCE: "Hierarchical Cluster Analysis" by V. J. de Ghett, 1978, *Quantitative Ethology*, ed. by P. W. Colgan. Reprinted by permission of John Wiley & Sons, Inc.

There are several ways to find clusters in this matrix. Two of them are called **single-link** or **closest-neighbor analysis** and **complete-link** or **farthest-neighbor analysis** (there are others, but I won't go into them here). In single-link clustering, we use only the numbers adjacent to the diagonal: 2, 4, 2, 5,

6, 1, 4. The two clustering solutions (again, done with *Anthropac*) are shown in figure 21.7.

a. SINGLE LINK								b. COMPLETE LINK										
	1	3	7	9	14	20	21	25		1	3	7	9	14	20	21	25	
Level									Level									
1						x	x	x	1							x	x	x
2	x	x	x				x	x	2	x	x	x				x	x	x
4	x	x	x	x	x	x			5	x	x	x	x	x		x	x	x
5	x	x	x	x	x	x	x	x	7	x	x	x	x	x	x	x	x	x
6	x	x	x	x	x	x	x	x	13	x	x	x	x	x	x	x	x	x
									24	x	x	x	x	x	x	x	x	x

Figure 21.7. Cluster analysis of data in table 21.27.

In the single-link solution, the two closest neighbors are 20 and 21. They are exactly one unit of distance apart, and there is a 1 adjacent to the diagonal of the original matrix where 20 and 21 come together. In figure 21.7a, 20 and 21 are shown joined at level 1. The numbers 1,3 and the numbers 7,9 are the next closest neighbors. They are both two units apart. Figure 21.7a shows them joined at level 2.

Once a pair is joined, it is considered a unit. The pairs 1,3 and 7,9 are joined together at level 4 because they are four units apart (the nearest neighbor to the pair 1,3 is 7, which is four units from 3). The pair 21,25 are also four units apart. However, 20,21 are already joined, so 25 joins this pair at level 4. The connections are built up to form a tree.

Figure 21.7b shows the complete-link (or farthest neighbor) clustering solution for the data in table 21.29. In complete-link clustering, all the numbers in table 21.29 are used. Once again, the pair 20,21 is joined at level 1 because the pair is just one unit apart. The pairs 1,3 and 7,9 join at level 2.

At this point, the complete-link and single-link solutions are identical. At the next level, though, things change. The neighbors of 20,21 are 14 and 25. The farthest neighbor from 14 to 20,21 is 21. The distance is seven units. The farthest neighbor from 25 to 20,21 is 20. The distance is five units. Since five is less than seven, 25 joins 20,21 at level 5. But the two pairs 1,3 and 7,9 are not joined at this level.

The only number not yet joined to some other number is 14. It is compared to its farthest neighbors in the adjacent clusters: 14 is 11 units away from 25 (which is now part of the 20,21,25 cluster) and it's seven units away from the 7,9 cluster. So, at level 7, 14 is joined to 7,9. The same game is played out with all the clusters to form the tree in figure 21.7b.

Lambros Comitas and I used DFA in our study of two groups of people in Athens, Greece: those who had returned from having spent at least 5 years in West Germany as labor migrants and those who had never been out of Greece. We were trying to understand how the experience abroad might have affected the attitudes of Greek men and women about traditional gender roles (Bernard and Comitas 1978). Our sample consisted of 400 persons: 100 male migrants, 100 female migrants, 100 male nonmigrants, and 100 female nonmigrants. Using DFA, we were able to predict with 70% accuracy whether an informant had been a migrant on the basis of just five variables.

There are some things you need to be careful about in using DFA, however. Notice that our sample in the Athens study consisted of half migrants and half nonmigrants. That was because we used a disproportionate, stratified sampling design to ensure adequate representation of returned migrants in the study. Given our sample, we could have guessed whether one of our informants was a migrant with 50% accuracy, without any information about the informant at all.

Now, only a very small fraction of the population of Athens consists of former long-term labor migrants to West Germany. The chances of stopping an Athenian at random on the street and grabbing one of those returned labor migrants was less than 5% in 1977 when we did the study.

Suppose that, armed with the results of the DFA that Comitas and I did, I asked random Athenians five questions, the answers to which allow me to predict 70% of the time whether any respondent had been a long-term labor migrant to West Germany. No matter what the answers were to those questions, I'd be better off predicting that the random Athenian was *not* a returned migrant. I'd be right more than 95% of the time.

Furthermore, why not just ask the random survey respondent straight out: "Are you a returned long-term labor migrant from West Germany?" With such an innocuous question, presumably I'd have gotten a correct answer at least as often as our 70% prediction based on knowing five pieces of information.

DFA is a powerful classification device, but it is not really a prediction device. Still, many problems (like the one Comitas and I studied) are essentially about understanding things so you can classify them correctly. Livingstone and Lunt (1993) surveyed 217 people in Oxford, England, and divided them into six types, based on whether or not people were in debt, whether or not people had savings, and people who were living exactly within their income (with neither savings nor debt). DFA, using a variety of variables (age, class, education, income, expenses, attitudes toward debt, etc.) correctly classified almost 95% of the cases into one of the six groups that Livingstone and Lunt had identified.

Gans and Wood (1985) used DFA technique for classifying Samoan women as “traditional” or “modern” with respect to their ideal family size. If women stated that they wanted three or fewer children, Gans and Wood placed them in a category they labeled “modern.” Women who said they wanted four or more children were labeled “traditional.” DFA showed that just six of the many variables that Gans and Wood had collected allowed them to classify correctly which category a woman belonged to in 75% of all cases. The variables were such things as age, owning a car, level of education, etc.

It would have been ridiculous for Gans and Wood to have asked women straight out: “Are you traditional or modern when it comes to the number of children you’d like?” DFA (combined with on-the-ground ethnography) gave them a good picture of the variables that go into Samoan women’s desired family size.

Similarly, Comitas and I were able to describe the attitudinal components of gender role changes by using DFA, and our prediction rate of 70% was significantly better than the 50% we’d have gotten by chance, given our sampling design. If you’re careful about how you interpret the results of a discriminant function analysis, it can be a really important addition to your statistical tool kit.

And Finally . . .

In a world of thousands of variables and millions of combinations of variables, how do you decide what to test? There is no magic formula. My advice is to follow every hunch you get. Some researchers insist that you have a good theoretical reason for including variables in your design and that you have a theory-driven reason to test for relations among variables once you have data. They point out that anyone can make up an explanation for any relation or lack of relation after seeing a table of data or a correlation coefficient.

This is very good advice, but I think it’s a bit too restrictive, for three reasons:

1. I think that data analysis should be lots of fun, and it can’t be unless it’s based on following your hunches. Most relations are easy to explain, and peculiar relations beg for theories to explain them. You just have to be very careful not to conjure up support for every statistically significant relation, merely because it happens to turn up. There is a delicate balance between being clever enough to explain an unexpected finding and just plain reaching too far. As usual, there is no substitute for thinking hard about your data.
2. It is really up to you during research design to be as clever as you can in thinking up variables to test. You’re entitled to include some variables in your research

just because you think they might come in handy. Of course, you can overdo it. There is nothing more tedious than an interview that drones on for hours without any obvious point other than that the researcher is gathering data on as many variables as possible.

3. The source of ideas has no necessary effect on their usefulness. You can get ideas from an existing theory or from browsing through data tables—or from talking about research problems with your friends. The important thing is not *how* you get a hunch, it's *whether you can test* your hunches and create plausible explanations for whatever findings come out of those tests. If others disagree with your explanations, then let them demonstrate that you are wrong, either by reanalyzing your data or by producing new data. But stumbling onto a significant relation between some variables does nothing to invalidate the relation.

So, when you design your research, try to think about the kinds of variables that might be useful in testing your hunches. Use the principles in chapter 5 and consider internal state variables (e.g., attitudes, values, beliefs); external state variables (e.g., age, height, gender, race, health status, occupation, wealth status); physical and cultural environmental variables (e.g., rainfall, socioeconomic class of a neighborhood); and time or space variables (Have attitudes changed over time? Do the people in one community behave differently from those in another otherwise similar community?).

In applied research, important variables are the ones that let you target a policy—that is, focus intervention efforts on subpopulations of interest (the rural elderly, victims of violent crime, overachieving third graders, etc.)—or that are more amenable to policy manipulation (knowledge is far more manipulable than attitudes or behavior, for example). No matter what the purposes of your research, or how you design it, the two principle rules of data analysis are:

1. If you have an idea, test it.
2. You can't test it if you don't have data on it.

So, don't be afraid to play and have a good time with data analysis. If you hang around people who use complex statistical tools in their research, you'll hear them talking about “massaging” their data, “teasing out signals” from their data, and “separating the signals from the noise.” These are not the sorts of phrases used by people who are bored with what they're doing.

Enjoy.

APPENDIX A

Table of Random Numbers

10097	32533	76520	13586	34673	54876	80959	09117	39292	74945
37542	04805	64894	74296	24805	24037	20636	10402	00822	91665
08422	68953	19645	09303	23209	02560	15953	34764	35080	33606
99019	02529	09376	70715	28311	31165	88676	74397	04436	27659
12807	99970	80157	36147	64032	36653	98951	16877	12171	76833
66065	74717	34072	76850	36697	36170	65813	39885	11199	29170
31060	10805	45571	82406	35303	42614	86799	07439	23403	09732
85269	77602	02051	65692	68665	74818	73053	85247	18623	88579
63573	32135	05325	47048	90553	57548	28468	28709	83491	25624
73796	45753	03529	64778	35808	34282	60935	20344	35273	88435
98520	17767	14905	68607	22109	40558	60970	93433	50500	73998
11805	05431	39808	27732	50725	68248	29405	24201	52775	67851
83452	99634	06288	98083	13746	70078	18475	40610	68711	77817
88685	40200	86507	58401	36766	67951	90364	76493	29609	11062
99594	67348	87517	64969	91826	08928	93785	61368	23478	34113
65481	17674	17468	50950	58047	76974	73039	57186	40218	16544
80124	35635	17727	08015	45318	22374	21115	78253	14385	53763
74350	99817	77402	77214	43236	00210	45521	64237	96286	02655
69916	26803	66252	29148	36936	87203	76621	13990	94400	56418
09893	20505	14225	68514	46427	56788	96297	78822	54382	14598
91499	14523	68479	27686	46162	83554	94750	89923	37089	20048
80336	94598	26940	36858	70297	34135	53140	33340	42050	82341
44104	81949	85157	47954	32979	26575	57600	40881	22222	06413
12550	73742	11100	02040	12860	74697	96644	89439	28707	25815
63606	49329	16505	34484	40219	52563	43651	77082	07207	31790
61196	90446	26457	47774	51924	33729	65394	59593	42582	60527
15474	45266	95270	79953	59367	83848	82396	10118	33211	59466
94557	28573	67897	54387	54622	44431	91190	42592	92927	45973
42481	16213	97344	08721	16868	48767	03071	12059	25701	46670
23523	78317	73208	89837	68935	91416	26252	29663	05522	82562
04493	52494	75246	33824	45862	51025	61962	79335	65337	12472
00549	97654	64051	88159	96119	63896	54692	82391	23287	29529
35963	15307	26898	09354	33351	35462	77974	50024	90103	39333
59808	08391	45427	26842	83609	49700	13021	24892	78565	20106
46058	85236	01390	92286	77281	44077	93910	83647	70617	42941

32179	00597	87379	25241	05567	07007	86743	17157	85394	11838
69234	61406	20117	45204	15956	60000	18743	92423	97118	96338
19565	41430	01758	75379	40419	21585	66674	36806	84962	85207
45155	14938	19476	07246	43667	94543	59047	90033	20826	69541
94864	31994	36168	10851	34888	81553	01540	35456	05014	51176
98086	24826	45240	28404	44999	08896	39094	73407	35441	31880
33185	16232	41941	50949	89435	48581	88695	41994	37548	73043
80951	00406	96382	70774	20151	23387	25016	25298	94624	61171
79752	49140	71961	28296	69861	02591	74852	20539	00387	59579
18633	32537	98145	06571	31010	24674	05455	61427	77938	91936
74029	43902	77557	32270	97790	17119	52527	58021	80814	51748
54178	45611	80993	37143	05335	12969	56127	19255	36040	90324
11664	49883	52079	84827	59381	71539	09973	33440	88461	23356
48324	77928	31249	64710	02295	36870	32307	57546	15020	09994
69074	94138	87637	91976	35584	04401	10518	21615	01848	76938
90089	90249	62196	53754	61007	39513	71877	19088	94091	97084
70413	74646	24580	74929	94902	71143	01816	06557	74936	44506
17022	85475	76454	97145	31850	33650	75223	90607	15520	39823
24906	46977	78868	59973	61110	13047	84302	15982	72731	82300
50222	97585	15161	11327	66712	76500	81055	43716	93343	02797
60291	56491	75093	71017	92139	21562	67305	33066	60719	20033
31485	66220	71939	23182	44059	00289	17996	05268	97659	02611
16551	13457	83006	43096	71235	29381	93168	46668	30723	29437
90831	40282	48952	90899	87567	14411	31483	78232	52117	57484
19195	94881	99625	59598	33330	34405	45601	39005	65170	48419
06056	81764	46911	33370	35719	30207	61967	08086	40073	75215
46044	94342	04346	25157	73062	41921	82742	70481	83376	28856
03690	95581	83895	32069	94196	93097	97900	79905	79610	68639
23532	45828	02575	70187	64732	95799	20005	44543	08965	58907
81365	88745	79117	66599	32463	76925	70223	80849	48500	92536
57660	57584	14276	10166	82132	61861	63597	91025	76338	06878
13619	18065	33262	41774	33145	69671	14920	62061	42352	61546
07155	33924	34103	48785	28604	75023	46564	44875	07478	61678
19705	73768	44407	66609	00883	56229	50882	76601	50403	18003
04233	69951	33035	72878	61494	38754	63112	34005	82115	72073
79786	96081	42535	47848	84053	38522	55756	20382	67816	84693
76421	34950	98800	04822	57743	40616	73751	36521	34591	68549
28120	11330	46035	36097	93141	90483	83329	51529	94974	86242
45012	95348	64843	44570	26086	57925	52060	86496	44979	45833
45251	99242	98656	72488	35515	08968	46711	56846	29418	15329

97318	06337	19410	09936	28536	08458	90982	66566	30286	27797
55895	62683	25132	51771	70516	05063	69361	75727	48522	89141
80181	03112	21819	10421	35725	92004	36822	18679	51605	48064
39423	21649	18389	01344	36548	07702	85187	75037	89625	39524
37040	87608	46311	03712	42044	33852	52206	86204	99714	82241
72664	17872	02627	65809	17307	97355	60006	18166	51375	79461
71584	11935	87348	22204	93483	37555	31381	23640	31469	92988
87697	30854	25509	22665	31581	12507	53679	26381	48023	47916
73663	27869	40208	40672	83210	48573	22406	46286	46987	12017
51544	01914	17431	97024	09620	54225	44529	90758	11151	98314
82670	82296	96903	45286	85145	60329	27682	64892	75961	19800
30051	16942	17241	93593	75336	48698	48564	76832	29214	84972
23338	01489	39942	06609	14070	07351	28226	51996	31244	10725
08739	21034	57145	25526	58145	72334	87799	95132	70300	88277
76383	52236	07587	14161	82994	22829	72713	70265	88650	56335
05933	81888	32534	56269	12889	05092	84159	40971	46430	86981
10347	07364	51963	31851	45463	41635	10195	18961	17515	34021
36102	55172	25170	81955	25621	25030	19781	48300	79319	34377
70791	56165	64310	28625	26760	82203	26535	99580	77676	91021
88525	67427	59554	42220	27202	18827	33362	90584	99516	72258
41221	71024	99746	77782	53452	52851	35104	20732	16072	72468
40771	10858	31707	46962	71427	85412	49561	93011	64079	38527
09913	14509	46399	82692	05526	19955	02385	85686	62040	39386
00420	06149	06188	72365	12603	83142	98814	66265	98583	93424
90748	19314	55032	64625	47855	32726	69744	54536	16494	33623

SOURCE: RAND Corporation, *A Million Random Digits with 100,000 Normal Deviates*. New York: Free Press, 1965, 1985. Reprinted by permission.

APPENDIX B

Table of Areas under a Normal Curve

(A)	(B)	(C)	(A)	(B)	(C)	(A)	(B)	(C)
z	<i>area between mean and z</i>	<i>area beyond z</i>	z	<i>area between mean and z</i>	<i>area beyond z</i>	z	<i>area between mean and z</i>	<i>area beyond z</i>
0.00	.0000	.5000	0.37	.1554	.3446	0.80	.2881	.2119
0.01	.0040	.4960	0.41	.1591	.3409	0.81	.2910	.2090
0.02	.0080	.4920	0.42	.1628	.3372	0.82	.2939	.2061
0.03	.0120	.4880	0.43	.1664	.3336	0.83	.2967	.2033
0.04	.0160	.4840	0.44	.1700	.3300	0.84	.2995	.2005
0.05	.0199	.4801	0.45	.1736	.3264	0.85	.3023	.1977
0.06	.0239	.4761	0.46	.1772	.3228	0.86	.3051	.1949
0.07	.0279	.4721	0.47	.1808	.3192	0.87	.3078	.1922
0.08	.0319	.4681	0.48	.1844	.3156	0.88	.3106	.1894
0.09	.0359	.4641	0.49	.1879	.3121	0.89	.3133	.1867
0.10	.0398	.4602	0.50	.1915	.3085	0.90	.3159	.1841
0.11	.0438	.4562	0.51	.1950	.3050	0.91	.3186	.1814
0.12	.0478	.4522	0.52	.1985	.3015	0.92	.3212	.1788
0.13	.0517	.4483	0.53	.2019	.2981	0.93	.3238	.1762
0.14	.0557	.4443	0.54	.2054	.2946	0.94	.3264	.1736
0.15	.0596	.4404	0.55	.2088	.2912	0.95	.3289	.1711
0.16	.0636	.4364	0.56	.2123	.2877	0.96	.3315	.1685
0.17	.0675	.4325	0.57	.2157	.2843	0.97	.3340	.1660
0.18	.0714	.4286	0.58	.2190	.2810	0.98	.3365	.1635
0.19	.0753	.4247	0.59	.2224	.2776	0.99	.3389	.1611
0.20	.0793	.4207	0.60	.2257	.2743	1.00	.3413	.1587
0.21	.0832	.4168	0.61	.2291	.2709	1.01	.3438	.1562
0.22	.0871	.4129	0.62	.2324	.2676	1.02	.3461	.1539
0.23	.0910	.4090	0.63	.2357	.2643	1.03	.3485	.1515
0.24	.0948	.4052	0.64	.2389	.2611	1.04	.3508	.1492
0.25	.0987	.4013	0.65	.2422	.2578	1.05	.3531	.1469
0.26	.1026	.3974	0.66	.2454	.2546	1.06	.3554	.1446
0.27	.1064	.3936	0.67	.2486	.2514	1.07	.3577	.1423
0.28	.1103	.3897	0.68	.2517	.2483	1.08	.3599	.1401
0.29	.1141	.3859	0.69	.2549	.2451	1.09	.3621	.1379
0.30	.1179	.3821	0.70	.2580	.2420	1.10	.3643	.1357
0.31	.1217	.3783	0.71	.2611	.2389	1.11	.3665	.1335
0.32	.1255	.3745	0.72	.2642	.2358	1.12	.3686	.1314
0.33	.1293	.3707	0.73	.2673	.2327	1.13	.3708	.1292
0.34	.1331	.3669	0.74	.2704	.2296	1.14	.3729	.1271
0.35	.1368	.3632	0.75	.2734	.2266	1.15	.3749	.1251
0.36	.1406	.3594	0.76	.2764	.2236	1.16	.3770	.1230
0.37	.1443	.3557	0.77	.2794	.2206	1.17	.3790	.1210
0.38	.1480	.3520	0.78	.2823	.2177	1.18	.3810	.1190
0.39	.1517	.3483	0.79	.2852	.2148	1.19	.3830	.1170

(A)	(B)	(C)	(A)	(B)	(C)	(A)	(B)	(C)
<i>z</i>	<i>area between mean and z</i>	<i>area beyond z</i>	<i>z</i>	<i>area between mean and z</i>	<i>area beyond z</i>	<i>z</i>	<i>area between mean and z</i>	<i>area beyond z</i>
1.20	.3849	.1151	1.61	.4463	.0537	2.02	.4783	.0217
1.21	.3869	.1131	1.62	.4474	.0526	2.03	.4788	.0212
1.22	.3888	.1112	1.63	.4484	.0516	2.04	.4793	.0207
1.23	.3907	.1093	1.64	.4495	.0505	2.05	.4798	.0202
1.24	.3925	.1075	1.65	.4505	.0495	2.06	.4803	.0197
1.25	.3944	.1056	1.66	.4515	.0485	2.07	.4808	.0192
1.26	.3962	.1038	1.67	.4525	.0475	2.08	.4812	.0188
1.27	.3980	.1020	1.68	.4535	.0465	2.09	.4817	.0183
1.28	.3997	.1003	1.69	.4545	.0455	2.10	.4821	.0179
1.29	.4015	.0985	1.70	.4554	.0446	2.11	.4826	.0174
1.30	.4032	.0968	1.71	.4564	.0436	2.12	.4830	.0170
1.31	.4049	.0951	1.72	.4573	.0427	2.13	.4834	.0166
1.32	.4066	.0934	1.73	.4582	.0418	2.14	.4838	.0162
1.33	.4082	.0918	1.74	.4591	.0409	2.15	.4842	.0158
1.34	.4099	.0901	1.75	.4599	.0401	2.16	.4846	.0154
1.35	.4115	.0885	1.76	.4608	.0392	2.17	.4850	.0150
1.36	.4131	.0869	1.77	.4616	.0384	2.18	.4854	.0146
1.37	.4147	.0853	1.78	.4625	.0375	2.19	.4857	.0143
1.38	.4162	.0838	1.79	.4633	.0367	2.20	.4861	.0139
1.39	.4177	.0823	1.80	.4641	.0359	2.21	.4864	.0136
1.40	.4192	.0808	1.81	.4649	.0351	2.22	.4868	.0132
1.41	.4207	.0793	1.82	.4656	.0344	2.23	.4871	.0129
1.42	.4222	.0778	1.83	.4664	.0336	2.24	.4875	.0125
1.43	.4236	.0764	1.84	.4671	.0329	2.25	.4878	.0122
1.44	.4251	.0749	1.85	.4678	.0322	2.26	.4881	.0119
1.45	.4265	.0735	1.86	.4686	.0314	2.27	.4884	.0116
1.46	.4279	.0721	1.87	.4693	.0307	2.28	.4887	.0113
1.47	.4292	.0708	1.88	.4699	.0301	2.29	.4890	.0110
1.48	.4306	.0694	1.89	.4706	.0294	2.30	.4893	.0107
1.49	.4319	.0681	1.90	.4713	.0287	2.31	.4896	.0104
1.50	.4332	.0668	1.91	.4719	.0281	2.32	.4898	.0102
1.51	.4345	.0655	1.92	.4726	.0274	2.33	.4901	.0099
1.52	.4357	.0643	1.93	.4732	.0268	2.34	.4904	.0096
1.53	.4370	.0630	1.94	.4738	.0262	2.35	.4906	.0094
1.54	.4382	.0618	1.95	.4744	.0256	2.36	.4909	.0091
1.55	.4394	.0606	1.96	.4750	.0250	2.37	.4911	.0089
1.56	.4406	.0594	1.97	.4756	.0244	2.38	.4913	.0087
1.57	.4418	.0582	1.98	.4761	.0239	2.39	.4916	.0084
1.58	.4429	.0571	1.99	.4767	.0233	2.40	.4918	.0082
1.59	.4441	.0559	2.00	.4772	.0228	2.41	.4920	.0080
1.60	.4452	.0548	2.01	.4778	.0222	2.42	.4922	.0078
2.43	.4925	.0075	2.74	.4969	.0031	3.05	.4989	.0011
2.44	.4927	.0073	2.75	.4970	.0030	3.06	.4989	.0011

(A)	(B)	(C)	(A)	(B)	(C)	(A)	(B)	(C)
z	area between mean and z	area beyond z	z	area between mean and z	area beyond z	z	area between mean and z	area beyond z
2.45	.4929	.0071	2.76	.4971	.0029	3.07	.4989	.0011
2.46	.4931	.0069	2.77	.4972	.0028	3.08	.4990	.0010
2.47	.4932	.0068	2.78	.4973	.0027	3.09	.4990	.0010
2.48	.4934	.0066	2.79	.4974	.0026	3.10	.4990	.0010
2.49	.4936	.0064	2.80	.4974	.0026	3.11	.4991	.0009
2.50	.4938	.0062	2.81	.4975	.0025	3.12	.4991	.0009
2.51	.4940	.0060	2.82	.4976	.0024	3.13	.4991	.0009
2.52	.4941	.0059	2.83	.4977	.0023	3.14	.4992	.0008
2.53	.4943	.0057	2.84	.4977	.0023	3.15	.4992	.0008
2.54	.4945	.0055	2.85	.4978	.0022	3.16	.4992	.0008
2.55	.4946	.0054	2.86	.4979	.0021	3.17	.4992	.0008
2.56	.4948	.0052	2.87	.4979	.0021	3.18	.4993	.0007
2.57	.4949	.0051	2.88	.4980	.0020	3.19	.4993	.0007
2.58	.4951	.0049	2.89	.4981	.0019	3.20	.4993	.0007
2.59	.4952	.0048	2.90	.4981	.0019	3.21	.4993	.0007
2.60	.4953	.0047	2.91	.4982	.0018	3.22	.4994	.0006
2.61	.4955	.0045	2.92	.4982	.0018	3.23	.4994	.0006
2.62	.4956	.0044	2.93	.4983	.0017	3.24	.4994	.0006
2.63	.4957	.0043	2.94	.4984	.0016	3.25	.4994	.0006
2.64	.4959	.0041	2.95	.4984	.0016	3.30	.4995	.0005
2.65	.4960	.0040	2.96	.4985	.0015	3.35	.4996	.0004
2.66	.4961	.0039	2.97	.4985	.0015	3.40	.4997	.0003
2.67	.4962	.0038	2.98	.4986	.0014	3.45	.4997	.0003
2.68	.4963	.0037	2.99	.4986	.0014	3.50	.4998	.0002
2.69	.4964	.0036	3.00	.4987	.0013	3.60	.4998	.0002
2.70	.4965	.0035	3.01	.4987	.0013	3.70	.4999	.0001
2.71	.4966	.0034	3.02	.4987	.0013	3.80	.4999	.0001
2.72	.4967	.0033	3.03	.4988	.0012	3.90	.49995	.00005
2.73	.4968	.0032	3.04	.4988	.0012	4.00	.49997	.00003

SOURCE: *Statistical Tables and Formulas*, Table 1, p. 3. New York: John Wiley, 1981. Reprinted by permission of A. Hald.

APPENDIX C

Student's *t* Distribution

<i>Level of significance for one-tailed test</i>						
	.10	.05	.025	.01	.005	.0005
<i>Level of significance for two-tailed test</i>						
<i>df</i>	.20	.10	.05	.02	.01	.001
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.941
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.859
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.405
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.767
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373
∞	1.282	1.645	1.960	2.326	2.567	3.291

SOURCE: *Statistical Tables for Biological, Agricultural and Medical Research*, R. A. Fisher and F. Yates, Pearson Education Limited. © 1963 R. A. Fisher and F. Yates. Printed and published in Great Britain by Oliver & Boyd Ltd., Edinburgh. Reprinted by permission.

APPENDIX D
Chi-Square Distribution Table

<i>df</i>	<i>Probabilities</i>					
	<i>.99</i>	<i>.95</i>	<i>.90</i>	<i>.80</i>	<i>.70</i>	<i>.50</i>
1	.000157	.00393	.0158	.0642	.148	.455
2	.0201	.103	.211	.446	.713	1.386
3	.115	.352	.584	1.005	1.424	2.366
4	.297	.711	1.064	1.649	2.195	3.357
5	.554	1.145	1.610	2.343	3.000	4.351
6	.872	1.635	2.204	3.070	3.828	5.348
7	1.239	2.167	2.833	3.822	4.671	6.346
8	1.646	2.733	3.490	4.594	5.527	7.344
9	2.088	3.325	4.168	5.380	6.393	8.343
10	2.558	3.940	4.865	6.179	7.267	9.342
11	3.053	4.575	5.578	6.989	8.148	10.341
12	3.571	5.226	6.304	7.807	9.034	11.340
13	4.107	5.892	7.042	8.634	9.926	12.340
14	4.660	6.571	7.790	9.467	10.821	13.339
15	5.229	7.261	8.547	10.307	11.721	14.339
16	5.812	7.962	9.312	11.152	12.624	15.338
17	6.408	8.672	10.085	12.002	13.531	16.338
18	7.015	9.390	10.865	12.857	14.440	17.338
19	7.633	10.117	11.651	13.716	15.352	18.338
20	8.260	10.851	12.443	14.578	16.266	19.337
21	8.897	11.591	13.240	15.445	17.182	20.337
22	9.542	12.338	14.041	16.314	18.101	21.337
23	10.196	13.091	14.848	17.187	19.021	22.337
24	10.865	13.848	15.659	18.062	19.943	23.337
25	11.524	14.611	16.473	18.940	20.867	24.337
26	12.198	15.379	17.292	19.820	21.792	25.336
27	12.879	16.151	18.114	20.703	22.719	26.336
28	13.565	16.928	18.939	21.588	23.647	27.336
29	14.256	17.708	19.768	22.475	24.577	28.336
30	14.953	18.493	20.599	23.364	25.508	29.336

<i>df</i>	<i>Probabilities</i>						
	<i>.30</i>	<i>.20</i>	<i>.10</i>	<i>.05</i>	<i>.025</i>	<i>.01</i>	<i>.001</i>
1	1.074	1.642	2.706	3.841	5.024	6.635	10.827
2	2.408	3.219	4.605	5.991	7.378	9.210	13.815
3	3.665	4.624	6.251	7.815	9.348	11.345	16.268
4	4.878	5.989	7.779	9.488	11.143	13.277	18.465
5	6.064	7.289	9.236	11.070	12.832	15.086	20.517
6	7.231	8.588	10.645	12.592	14.449	16.812	22.457
7	8.383	9.803	12.017	14.067	16.013	18.475	24.322
8	9.524	11.030	13.362	15.507	17.535	20.090	26.125
9	10.656	12.242	14.684	16.919	19.023	21.666	27.877
10	11.781	13.442	15.987	18.307	20.483	23.209	29.588
11	12.899	14.631	17.275	19.675	21.920	24.725	31.264
12	14.011	15.812	18.549	21.026	23.337	26.217	32.909
13	15.119	16.985	19.812	22.362	24.736	27.688	34.528
14	16.222	18.151	21.064	23.685	26.119	29.141	36.123
15	17.322	19.311	22.307	24.996	27.488	30.578	37.697
16	18.418	20.465	23.542	26.296	28.845	32.000	39.252
17	19.511	21.615	24.769	27.587	30.191	33.409	40.790
18	20.601	22.760	25.989	28.869	31.526	34.805	42.312
19	21.689	23.900	27.204	30.144	32.852	36.191	43.820
20	22.775	25.038	28.412	31.410	34.170	37.566	45.315
21	23.858	26.171	29.615	32.671	35.479	38.932	46.797
22	24.939	27.301	30.813	33.924	36.781	40.289	48.268
23	26.018	28.429	32.007	35.172	38.076	41.638	49.728
24	27.096	29.553	33.196	36.415	39.364	42.980	51.179
25	28.172	30.675	34.382	37.652	40.646	44.314	52.620
26	29.246	31.795	35.563	38.885	41.923	45.642	54.052
27	30.319	32.912	36.741	40.113	43.194	46.963	55.476
28	31.391	34.027	37.916	41.337	44.461	48.278	56.893
29	32.461	35.139	39.087	42.557	45.722	49.588	58.302
30	33.530	36.250	40.256	43.773	46.979	50.892	59.703

SOURCE: *Statistical Tables for Biological, Agricultural and Medical Research*, R. A. Fisher and F. Yates, Pearson Education Limited. © 1963 R. A. Fisher and F. Yates. Printed and published in Great Britain by Oliver & Boyd Ltd., Edinburgh. Reprinted by permission.

APPENDIX E
F Table for the .05 Level of Significance

$df_1 \backslash df_2$		<i>Numerator Degrees of Freedom</i>							
		1	2	3	4	5	6	8	10
Degrees of Freedom for the Denominator	1	161.4	199.5	215.7	224.6	230.2	234.0	238.9	241.9
	2	18.51	19.00	19.16	19.25	19.30	19.33	19.37	19.40
	3	10.13	9.55	9.28	9.12	9.01	8.94	8.85	8.79
	4	7.71	6.94	6.59	6.39	6.26	6.16	6.04	5.96
	5	6.61	5.79	5.41	5.19	5.05	4.95	4.82	4.74
	6	5.99	5.14	4.76	4.53	4.39	4.28	4.15	4.06
	7	5.59	4.74	4.35	4.12	3.97	3.87	3.73	3.64
	8	5.32	4.46	4.07	3.84	3.69	3.58	3.44	3.35
	9	5.12	4.26	3.86	3.63	3.48	3.37	3.23	3.14
	10	4.96	4.10	3.71	3.48	3.33	3.22	3.07	2.98
	11	4.84	3.98	3.59	3.36	3.20	3.09	2.95	2.85
	12	4.75	3.89	3.49	3.26	3.11	3.00	2.85	2.75
	13	4.67	3.81	3.41	3.18	3.03	2.92	2.77	2.67
	14	4.60	3.74	3.34	3.11	2.96	2.85	2.70	2.60
	15	4.54	3.68	3.29	3.06	2.90	2.79	2.64	2.54
	16	4.49	3.63	3.24	3.01	2.85	2.74	2.59	2.49
	17	4.45	3.59	3.20	2.96	2.81	2.70	2.55	2.45
	18	4.41	3.55	3.16	2.93	2.77	2.66	2.51	2.41
	19	4.38	3.52	3.13	2.90	2.74	2.63	2.48	2.38
	20	4.35	3.49	3.10	2.87	2.71	2.60	2.45	2.35
	21	4.32	3.47	3.07	2.84	2.68	2.57	2.42	2.32
	22	4.30	3.44	3.05	2.82	2.66	2.55	2.40	2.30
	23	4.28	3.42	3.03	2.80	2.64	2.53	2.37	2.27
	24	4.26	3.40	3.01	2.78	2.62	2.51	2.36	2.25
	25	4.24	3.39	2.99	2.76	2.60	2.49	2.34	2.24
	26	4.23	3.37	2.98	2.74	2.59	2.47	2.32	2.22
	27	4.21	3.35	2.96	2.73	2.57	2.46	2.31	2.20
	28	4.20	3.34	2.95	2.71	2.56	2.45	2.29	2.19
	29	4.18	3.33	2.93	2.70	2.55	2.43	2.28	2.18
	30	4.17	3.32	2.92	2.69	2.53	2.42	2.27	2.16
40	4.08	3.23	2.84	2.61	2.45	2.34	2.18	2.08	
60	4.00	3.15	2.76	2.53	2.37	2.25	2.10	1.99	
80	3.96	3.11	2.72	2.48	2.33	2.21	2.05	1.95	
120	3.92	3.07	2.68	2.45	2.29	2.17	2.02	1.91	
∞	3.84	3.00	2.60	2.37	2.21	2.10	1.94	1.83	

$\begin{matrix} df_1 \\ \backslash \\ df_2 \end{matrix}$		<i>Numerator Degrees of Freedom</i>							
		12	15	20	30	40	60	120	∞
Degrees of Freedom for the Denominator	1	243.9	245.9	248.0	250.1	251.1	252.2	253.3	254.3
	2	19.41	19.43	19.45	19.46	19.47	19.48	19.49	19.50
	3	8.74	8.70	8.66	8.62	8.59	8.57	8.55	8.53
	4	5.91	5.86	5.80	5.75	5.72	5.69	5.66	5.63
	5	4.68	4.62	4.56	4.50	4.46	4.43	4.40	4.36
	6	4.00	3.94	3.87	3.81	3.77	3.74	3.70	3.67
	7	3.57	3.51	3.44	3.38	3.34	3.30	3.27	3.23
	8	3.28	3.22	3.15	3.08	3.04	3.01	2.97	2.93
	9	3.07	3.01	2.94	2.86	2.83	2.79	2.75	2.71
	10	2.91	2.85	2.77	2.70	2.66	2.62	2.58	2.54
	11	2.79	2.72	2.65	2.57	2.53	2.49	2.45	2.40
	12	2.69	2.62	2.54	2.47	2.43	2.38	2.34	2.30
	13	2.60	2.53	2.46	2.38	2.34	2.30	2.25	2.21
	14	2.53	2.46	2.39	2.31	2.27	2.22	2.18	2.13
	15	2.48	2.40	2.33	2.25	2.20	2.16	2.11	2.07
	16	2.42	2.35	2.28	2.19	2.15	2.11	2.06	2.01
	17	2.38	2.31	2.23	2.15	2.10	2.06	2.01	1.96
	18	2.34	2.27	2.19	2.11	2.06	2.02	1.97	1.92
	19	2.31	2.23	2.16	2.07	2.03	1.98	1.93	1.88
	20	2.28	2.20	2.12	2.04	1.99	1.95	1.90	1.84
	21	2.25	2.18	2.10	2.01	1.96	1.92	1.87	1.81
	22	2.23	2.15	2.07	1.98	1.94	1.89	1.84	1.78
	23	2.20	2.13	2.05	1.96	1.91	1.86	1.81	1.76
	24	2.18	2.11	2.03	1.94	1.89	1.84	1.79	1.73
	25	2.16	2.09	2.01	1.92	1.87	1.82	1.77	1.71
	26	2.15	2.07	1.99	1.90	1.85	1.80	1.75	1.69
	27	2.13	2.06	1.97	1.88	1.84	1.79	1.73	1.67
	28	2.12	2.04	1.96	1.87	1.82	1.77	1.71	1.65
	29	2.10	2.03	1.94	1.85	1.81	1.75	1.70	1.64
	30	2.09	2.01	1.93	1.84	1.79	1.74	1.68	1.62
40	2.00	1.92	1.84	1.74	1.69	1.64	1.58	1.51	
60	1.92	1.84	1.75	1.65	1.59	1.53	1.47	1.39	
80	1.88	1.80	1.70	1.60	1.54	1.49	1.41	1.32	
120	1.83	1.75	1.66	1.55	1.50	1.43	1.35	1.25	
∞	1.75	1.67	1.57	1.46	1.39	1.32	1.22	1.00	

F Table for the .01 Level of Significance

$df_1 \backslash df_2$		Numerator Degrees of Freedom							
		1	2	3	4	5	6	8	10
Degrees of Freedom for the Denominator	1	4052	4999.5	5403	5625	5764	5859	5982	6056
	2	98.50	99.00	99.17	99.25	99.30	99.33	99.37	99.40
	3	34.12	30.82	29.46	28.71	28.24	27.91	27.49	27.23
	4	21.20	18.00	16.69	15.98	15.52	15.21	14.80	14.55
	5	16.26	13.27	12.06	11.39	10.97	10.67	10.29	10.05
	6	13.75	10.92	9.78	9.15	8.75	8.47	8.10	7.87
	7	12.25	9.55	8.45	7.85	7.46	7.19	6.84	6.62
	8	11.26	8.65	7.59	7.01	6.63	6.37	6.03	5.81
	9	10.56	8.02	6.99	6.42	6.06	5.80	5.47	5.26
	10	10.04	7.56	6.55	5.99	5.64	5.39	5.06	4.85
	11	9.65	7.21	6.22	5.67	5.32	5.07	4.74	4.54
	12	9.33	6.93	5.95	5.41	5.06	4.82	4.50	4.30
	13	9.07	6.70	5.74	5.21	4.86	4.62	4.30	4.10
	14	8.86	6.51	5.56	5.04	4.69	4.46	4.14	3.94
	15	8.68	6.36	5.42	4.89	4.56	4.32	4.00	3.80
	16	8.53	6.23	5.29	4.77	4.44	4.20	3.89	3.69
	17	8.40	6.11	5.18	4.67	4.34	4.10	3.79	3.59
	18	8.29	6.01	5.09	4.58	4.25	4.01	3.71	3.51
	19	8.18	5.93	5.01	4.50	4.17	3.94	3.63	3.43
	20	8.10	5.85	4.94	4.43	4.10	3.87	3.56	3.37
	21	8.02	5.78	4.87	4.37	4.04	3.81	3.51	3.31
	22	7.95	5.72	4.82	4.31	3.99	3.76	3.45	3.26
	23	7.88	5.66	4.76	4.26	3.94	3.71	3.41	3.21
	24	7.82	5.61	4.72	4.22	3.90	3.67	3.36	3.17
	25	7.77	5.57	4.68	4.18	3.85	3.63	3.32	3.13
	26	7.72	5.53	4.64	4.14	3.82	3.59	3.29	3.09
	27	7.68	5.49	4.60	4.11	3.78	3.56	3.26	3.06
	28	7.64	5.45	4.57	4.07	3.75	3.53	3.23	3.03
	29	7.60	5.42	4.54	4.04	3.73	3.50	3.20	3.00
	30	7.56	5.39	4.51	4.02	3.70	3.47	3.17	2.98
	40	7.31	5.18	4.31	3.83	3.51	3.29	2.99	2.80
	60	7.08	4.98	4.13	3.65	3.34	3.12	2.82	2.63
80	6.96	4.88	4.04	3.56	3.25	3.04	2.74	2.55	
120	6.85	4.79	3.95	3.48	3.17	2.96	2.66	2.47	
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.51	2.32	

$\begin{matrix} df_1 \\ \backslash \\ df_2 \end{matrix}$		Numerator Degrees of Freedom							
		12	15	20	30	40	60	120	∞
Degrees of Freedom for the Denominator	1	6106	6157	6209	6261	6287	6313	6339	6366
	2	99.42	99.43	99.45	99.47	99.47	99.48	99.49	99.50
	3	27.05	26.87	26.69	26.50	26.41	26.32	26.22	26.13
	4	14.37	14.20	14.02	13.84	13.75	13.65	13.56	13.46
	5	9.89	9.72	9.55	9.38	9.29	9.20	9.11	9.02
	6	7.72	7.56	7.40	7.23	7.14	7.06	6.97	6.86
	7	6.47	6.31	6.16	5.99	5.91	5.82	5.74	5.65
	8	5.67	5.52	5.36	5.20	5.12	5.03	4.95	4.86
	9	5.11	4.96	4.81	4.65	4.57	4.48	4.40	4.31
	10	4.71	4.56	4.41	4.25	4.17	4.08	4.00	3.91
	11	4.40	4.25	4.10	3.94	3.86	3.78	3.69	3.60
	12	4.16	4.01	3.86	3.70	3.62	3.54	3.45	3.36
	13	3.96	3.82	3.66	3.51	3.43	3.34	3.25	3.17
	14	3.80	3.66	3.51	3.35	3.27	3.18	3.09	3.00
	15	3.67	3.52	3.37	3.21	3.13	3.05	2.96	2.87
	16	3.55	3.41	3.26	3.10	3.02	2.93	2.84	2.75
	17	3.46	3.31	3.16	3.00	2.92	2.83	2.75	2.65
	18	3.37	3.23	3.08	2.92	2.84	2.75	2.66	2.57
	19	3.30	3.15	3.00	2.84	2.76	2.67	2.58	2.49
	20	3.23	3.09	2.94	2.78	2.69	2.61	2.52	2.42
	21	3.17	3.03	2.88	2.72	2.64	2.55	2.46	2.36
	22	3.12	2.98	2.83	2.67	2.58	2.50	2.40	2.31
	23	3.07	2.93	2.78	2.62	2.54	2.45	2.35	2.26
	24	3.03	2.89	2.74	2.58	2.49	2.40	2.31	2.21
	25	2.99	2.85	2.70	2.54	2.45	2.36	2.27	2.17
	26	2.96	2.81	2.66	2.50	2.42	2.33	2.23	2.13
	27	2.93	2.78	2.63	2.47	2.38	2.29	2.20	2.10
	28	2.90	2.75	2.60	2.44	2.35	2.26	2.17	2.06
	29	2.87	2.73	2.57	2.41	2.33	2.23	2.14	2.03
	30	2.84	2.70	2.55	2.39	2.30	2.21	2.11	2.01
40	2.66	2.52	2.37	2.20	2.11	2.02	1.92	1.80	
60	2.50	2.35	2.20	2.03	1.94	1.84	1.73	1.60	
80	2.41	2.28	2.11	1.94	1.84	1.75	1.63	1.49	
120	2.34	2.19	2.03	1.86	1.76	1.66	1.53	1.38	
∞	2.18	2.04	1.88	1.70	1.59	1.47	1.32	1.00	

SOURCE: C. M. Thompson, "Tables of the Percentage Points of the Inverted Beta (F) Distribution," *Biometrika* 33 (1943), pp. 73–88. Reprinted by permission of Oxford University Press.

APPENDIX F
Resources for Fieldworkers

<http://www.qualquant.net/training/resources.htm>

References

- Abel, M. H. 1997. Low birth weight and interactions between traditional risk factors. *Journal of Genetic Psychology* 158:443–56.
- Adams, K. 1993. Dowry as compensatory marriage payment. Unpublished manuscript.
- Addams, J. 1926. *Twenty years at Hull House*. New York: Macmillan.
- Agar, M. 1973. *Ripping and running*. New York: Academic Press.
- Agar, M. 1980a. Getting better quality stuff: Methodological competition in an interdisciplinary niche. *Urban Life* 9:34–50.
- Agar, M. 1980b. *The professional stranger*. New York: Academic Press.
- Agar, M. 1996. *The professional stranger: An informal introduction to ethnography*. 2d ed. San Diego: Academic Press.
- Agar, M., and J. R. Hobbs. 1985. How to grow schemata out of interviews. In *Directions in cognitive anthropology*, edited by J.W.D. Dougherty, 413–31. Urbana: University of Illinois Press.
- Ahern, K., and R. Le Brocque. 2005. Methodological issues in the effects of attrition: Simple solutions for social scientists. *Field Methods* 17:53–69.
- Ahrentzen, S., D. W. Levine, and W. Michelson. 1989. Space, time, and activity in the home: A gender analysis. *Journal of Environmental Psychology* 9:89–101.
- Aiello, J. R., and S. E. Jones. 1971. Field study of the proxemic behavior of young school children in three subcultural groups. *Journal of Personality and Social Psychology* 19:351–56.
- Akkerboom, H., and H. Schmeets. 1998. Survey data collection methodology for election missions: How professional form design helps to control a survey. *International Journal of Public Opinion Research* 10(3): 257–74.
- Aldridge, D. 2004. *Health, the individual, and integrated medicine: Revisiting an aesthetic of health care*. London: Jessica Kingsley Publishers.
- Alexander, J. 1976. A study of the cultural domain of “relatives.” *American Ethnologist* 3:17–38.
- Algase, D. L., B. Kupferschmid, C. A. Beel-Bates, and E.R.A. Beattie. 1997. Estimates of stability of daily wandering behavior among cognitively impaired long-term care residents. *Nursing Research* 46:172–78.
- Al-Krenawi A., and R. Wiesel-Lev. 1999. Attitudes toward and perceived psychosocial impact of female circumcision as practiced among the Bedouin-Arabs of the Negev. *Family Process* 38(4): 431–43.
- Allen, J. T., and G. Italie. 1954. *A concordance to Euripides*. Berkeley: University of California Press.
- Allport, F. H., and D. A. Hartman. 1931. The prediction of cultural change: A problem illustrated in studies by F. Stuart Chapin and A. L. Kroeber. In *Methods in social science*, edited by S. A. Rice, 307–52. Chicago: University of Chicago Press.
- Al-Nuaim, A. A., E. A. Bamgboye, K. A. Al-Rubeaan, and Y. Al-Mazrou. 1997. Overweight and

- obesity in Saudi Arabian adult population. Role of sociodemographic variables. *Journal of Community Health* 22:211–23.
- Altmann, J. 1974. Observational study of behavior: Sampling methods. *Behaviour* 49:227–67.
- Altorki, S., and C. Fawzi El-Solh, eds. 1988. *Arab women in the field: Studying your own society*. Syracuse, N.Y.: Syracuse University Press.
- American Statistical Association. 1974. Report on the ASA Conference on Surveys of Human Populations. *The American Statistician* 28:30–34.
- Anderson, A. 2003. *A place on the corner*. Chicago: University of Chicago Press.
- Anderson, A. B., and E. S. Anderson. 1983. People and the Palm Forest. (Contract 51–07–79–07 to John Ewel.) Final report to United States Dept. of Agriculture, Forest Service, Consortium for the Study of Man's Relationship with the Global Environment. Washington, D.C.: USDA. (Also available through NTIS.)
- Anderson, C. A. 1989. Temperature and aggression: Ubiquitous effects of heat on occurrence of human violence. *Psychological Bulletin* 106:74–96.
- Anderson, M., and M. Halpin, eds. 2000. In *Potlatch at Gitsegukla: William Beynon's 1945 field notebooks*, edited by W. Beynon, 3–52. Vancouver: UBC Press.
- Angrosino, M. V. 1997. *Opportunity House: Ethnographic stories of mental retardation*. Walnut Creek, Calif.: AltaMira.
- Aquilino, W. S. 1993. Effects of spouse presence during the interview on survey responses concerning marriage. *Public Opinion Quarterly* 57:358–76.
- Aquilino, W. S. 1994. Interview mode effects in surveys of drug and alcohol use. *Public Opinion Quarterly* 58:210–40.
- Aquilino, W. S. 1997. Privacy effects on self-reported drug use: Interactions with survey mode and respondent characteristics. In *The validity of self-reported drug use: Improving the accuracy of survey estimates*, edited by L. Harrison and A. Hughes, 383–415. Rockville, Md.: National Institute on Drug Abuse, Division of Epidemiology and Prevention Research.
- Aquilino, W. S., D. L. Wright, and A. J. Supple. 2000. Response effects due to bystander presence in CASI and paper-and-pencil surveys of drug use and alcohol use. *Substance Use and Misuse* 35(6–8): 845–867. Special issue: *Methodological Issues in the Measurement of Drug Use*.
- Archer, D., and L. Erlich. 1985. Weighing the evidence: A new method for research on restricted information. *Qualitative Sociology* 8:345–58.
- Ardener, S., ed. 1981. *Women and space: Ground rules and social maps*. New York: St. Martin's Press.
- Arens, W. 1986. *The original sin: Incest and its meaning*. New York: Oxford University Press.
- Aronson, E., and J. Mills. 1959. The effect of severity of initiation on liking for a group. *Journal of Abnormal and Social Psychology* 59:177–81.
- Arp, W., III, and K. Boeckelman. 1997. Religiosity: A source of black environmentalism and empowerment? *Journal of Black Studies* 28:255–67.
- Associated Press. 1985. 1 October.
- Atkinson, J. M., and J. Heritage, eds. 1984. *Structures of social action: Studies in conversation analysis*. New York: Cambridge University Press.
- Atzwanger, K., K. Grammer, K. Schäfer, and A. Schmitt, eds. 1997. *New aspects of human ethnology*. New York: Plenum Press.
- Aunger, R. 1992. Sources of variation in ethnographic interview data: The case of food avoidances in the Ituri forest, Zaire. Paper presented at the annual meeting of the American Anthropological Association, San Francisco, California.
- Aunger, R. 2004. *Reflexive ethnographic science*. Walnut Creek, Calif.: AltaMira.

- Auster, C. J., and S. C. Ohm. 2000. Masculinity and femininity in contemporary American society: A reevaluation using the Bex Sex-Role Inventory. *Sex Roles* 43:499–528.
- Axinn, W. G. 1991. The influence of interviewer sex on responses to sensitive questions in Nepal. *Social Science Research* 20:303–19.
- Axinn, W. G., T. E. Fricke, and A. Thornton. 1991. The microdemographic community-study approach. *Sociological Methods and Research* 20:187–217.
- Ayres, I. 1991. Fair driving: Gender and race discrimination in retail car negotiations. *Harvard Law Review* 104:817–72.
- Babbie, E. 1983. *The practice of social research*. 3rd ed. Belmont, Calif.: Wadsworth.
- Baker, J., M. Levy, and D. Grewal. 1992. An experimental approach to making retail store environmental decisions. *Journal of Retailing* 68:445–60.
- Baker, R. 1996a. PRA with street children in Nepal. *PLA Notes* 25:56–60. London: International Institute for Environment and Development.
- Baker, R., with C. Panter-Brick and A. Todd. 1996b. Methods used in research with street children in Nepal. *Childhood* 3:171–93.
- Baksh, M., C. G. Neumann, M. Paglissio, R. M. Trostle, and A.A.J. Jansen. 1994. The influence of reproductive status on rural Kenyan women's time use. *Social Science and Medicine* 39(3): 345–54.
- Bales, R. F. 1950. *Interaction process analysis. A method for the study of small groups*. Cambridge, Mass.: Addison-Wesley.
- Bales, R. F. 2001. *Social interaction systems: Theory and measurement*. Piscataway, N.J.: Transaction.
- Bales, R. F., and S. P. Cohen. 1979. *SYMLOG: A system for the multiple level observation of groups*. New York: The Free Press.
- Balicki, A. 1989. *The Netsilik Eskimo*. Prospect Heights, Ill.: Waveland.
- Barber, N. 1998. Ecological and psychosocial correlates of male homosexuality: A cross-cultural investigation. *Journal of Cross-Cultural Psychology* 29:387–401.
- Barker, J. C. 1999. *Danger, duty and disillusion: The worldview of Los Angeles police officers*. Prospect Heights, Ill.: Waveland.
- Barnes, B. 2004. For Nielsen, fixing old ratings system causes new static. *The Wall Street Journal*, 16 September:1, 8.
- Barnes, J. H., B. F. Banahan, III, and K. E. Fish. 1995. The response effect of question order in computer-administered questioning in the social sciences. *Social Science Computer Review* 13:47–63.
- Barroso, J. 1997. Reconstructing my life: Becoming a long-term survivor of AIDS. *Qualitative Health Research* 7:57–74.
- Barry, H., III, and A. Schlegel. 1980. *Cross-cultural samples and codes*. Pittsburgh: University of Pittsburgh Press.
- Bartlett, F. C. 1937. Psychological methods and anthropological problems. *Africa* 10:401–19.
- Basiotis, P. P., S. O. Welsh, F. J. Cronin, J. L. Kelsay, and W. Mertz. 1987. Number of days of food intake records required to estimate individual and group nutrient intakes with defined confidence. *Journal of Nutrition* 117:1638–41.
- Bassuk, E. L., and L. Rosenberg. 1988. Why does family homelessness occur? A case-control study. *American Journal of Public Health* 78:783–88.
- Baugh, J. 1983. *Black street speech: Its history, structure, and survival*. Austin: University of Texas Press.
- Beardsworth, A., and T. Keil. 1992. The vegetarian option: Varieties, conversions, motives and careers. *The Sociological Review* 40:253–93.

- Beautrais, A. L., P. R. Joyce, and R. T. Mulder. 1998. Youth suicide attempts: A social and demographic profile. *Australian & New Zealand Journal of Psychiatry* 32(3): 349–57.
- Bechky, B. A. 2003. Object lessons: Workplace artifacts as representations of occupational jurisdiction. *American Journal of Sociology* 109:720–52.
- Bechtel, R. B. 1977. *Enclosing behavior*. Stroudsburg, Penn.: Dowden, Hutchinson, & Ross.
- Beck, A. T., C. H. Ward, M. Mendelson, J. E. Mock, and J. Erbaugh. 1961. An inventory for measuring depression. *Archives of General Psychiatry* 4:561–71.
- Beck, C. T., and R. K. Gable 2000. Postpartum Depression Screening Scale: Development and psychometric testing. *Nursing Research* 49(5): 272–82.
- Beck, C. T., and R. K. Gable 2003. Postpartum Depression Screening Scale: Spanish Version. *Nursing Research* 52(5): 296–306.
- Becker, H. S., B. Geer, E. C. Hughs, and A. L. Strauss. 1961. *Boys in white. Student culture in medical school*. Chicago: University of Chicago Press.
- Beere, C. A. 1990. *Gender roles: A handbook of tests and measures*. New York: Greenwood Press.
- Begley, S., H. Fineman, and V. Church. 1992. The science of polling. *Newsweek*, 28 September: 38–39.
- Behar, R. 1996. *The vulnerable observer: Anthropology that breaks your heart*. Boston: Beacon Press.
- Belk, R. W., J. F. Sherry, Jr., and M. Wallendorf. 1988. A naturalistic inquiry into buyer and seller behavior at a swap meet. *Journal of Consumer Research* 14:449–69.
- Belli, R. F. 1998. The structure of autobiographical memory and the event history calendar: Potential improvements in the quality of retrospective reports in surveys. *Memory* 6:383–406.
- Belo, J. 1960. *Trance in Bali*. New York: Columbia University Press.
- Belter, R. W., J. A. McIntosh, A. J. Finch, Jr., and C. F. Saylor. 1988. Preschoolers' ability to differentiate levels of pain: Relative efficacy of three self-report measures. *Journal of Clinical Child Psychology* 17:329–35.
- Bem, S. L. 1974. The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology* 42:155–62.
- Bem, S. L. 1979. Theory and measurement of androgyny: A reply to the Pedhazur-Tetenbaum and Locksley-Colten critiques. *Journal of Personality and Social Psychology* 37:1047–54.
- Benedict, R. 1946. *The chrysanthemum and the sword*. Boston: Houghton Mifflin.
- Bentley, M. E., G. H. Peltó, W. L. Straus, D. A. Schumann, C. Adegbola, E. de la Pena, G. A. Oni, K. H. Brown, and S. L. Huffman. 1988. Rapid ethnographic assessment: Applications in a diarrhea management program. *Social Science and Medicine* 27:107–16.
- Berg, D. N., and K. K. Smith, eds. 1985. *Exploring clinical methods for social research*. Beverly Hills, Calif.: Sage.
- Berman, J. 1996. "The culture as it appears to the Indian himself": Boas, George Hunt, and the methods of ethnography. In *Volksgeist as method and ethic*, edited by G. W. Stocking, Jr., 215–56. Madison: University of Wisconsin Press.
- Bermant, G. 1982. Justifying social research in terms of social benefit. In *Ethical issues in social science research*, edited by T. L. Beauchamp et al., 125–43. Baltimore: The Johns Hopkins University Press.
- Bernard, H. R. 1965. Greek sponge boats in Florida. *Anthropological Quarterly* 38(2): 41–54.
- Bernard, H. R. 1967. Kalyrnian sponge diving. *Human Biology* 39:103–30.
- Bernard, H. R. 1974. Scientists and policymakers: A case study in the ethnography of communications. *Human Organization* 33:261–75.
- Bernard, H. R. 1987. Sponge fishing and technological change in Greece. In *Technology and*

- social change*, 2d ed., edited by H. R. Bernard and P. J. Pelto, 167–206. Prospect Heights, Ill.: Waveland.
- Bernard, H. R. 1992. Preserving language diversity. *Human Organization* 41:82–88.
- Bernard, H. R. 1993. Methods belong to all of us. In *Assessing cultural anthropology*, edited by R. Borofsky, 168–78. New York: McGraw Hill.
- Bernard, H. R. 1996. Language preservation and publishing. In *Indigenous literacies in the Americas*, edited by N. Hornberger, 139–56. Berlin: Mouton de Gruyter.
- Bernard, H. R., and S. Ashton-Vouyoucalos. 1976. Return migration to Greece. *Journal of the Steward Anthropological Society* 8:31–51.
- Bernard, H. R., and L. Comitas. 1978. Greek return migration. *Current Anthropology* 19:658–59.
- Bernard, H. R., and M. J. Evans. 1983. New microcomputer techniques for anthropologists. *Human Organization* 42:182–85.
- Bernard, H. R., and W. P. Handwerker. 1995. *Data analysis with MYSTAT*. New York: McGraw-Hill.
- Bernard, H. R., E. Johnsen, P. Killworth, and S. Robinson. 1989. Estimating the size of an average personal network and of an event population. In *The small world*, edited by M. Kochen, 159–75. Norwood, N.J.: Ablex.
- Bernard, H. R., and P. D. Killworth. 1973. On the social structure of an ocean-going research vessel and other important things. *Social Science Research* 2:145–84.
- Bernard, H. R., and P. D. Killworth. 1974. Scientists and crew. *Maritime Studies and Management* 2:112–25.
- Bernard, H. R., and P. D. Killworth. 1993. Sampling in time allocation research. *Ethnology* 32:211.
- Bernard, H. R., P. D. Killworth, M. J. Evans, C. McCarty, and G. A. Shelley. 1987. Measuring patterns of acquaintanceship cross-culturally. *Ethnology* 27:155–79.
- Bernard, H. R., P. D. Killworth, L. Sailer, and D. Kronenfeld. 1984. The problem of informant accuracy: The validity of retrospective data. *Annual Review of Anthropology* 13:495–517.
- Bernard, H. R., and P. J. Pelto. 1987. Technology and anthropological theory. In *Technology and social change*, 2d ed. rev., edited by H. R. Bernard and P. J. Pelto, 359–76. Prospect Heights, Ill.: Waveland.
- Bernard, H. R., G. Ryan, and S. Borgatti. 1999. Modeling some green behaviors: Why Americans do or don't recycle aluminum cans. Unpublished manuscript.
- Bernard, H. R., and J. Salinas Pedraza, eds. 1976. Otomí parables, folktales, and jokes. *International Journal of American Linguistics*, Native American Text Series, Vol. 1, No. 2. University of Chicago Press.
- Bernard, H. R., and J. Salinas Pedraza. 1989. *Native ethnography: A Mexican Indian describes his culture*. Newbury Park, Calif.: Sage.
- Berra, Y., and J. Garagiola. 1998. *The Yogi book : 'I really didn't say everything I said.'* New York: Workman Publishing.
- Berreman, G. D. 1962. *Behind many masks*. Ithaca, N.Y.: Society for Applied Anthropology.
- Berry, J. 1976. *Human ecology and cognitive style*. New York: Wiley.
- Bessinger, J. B., and P. H. Smith 1969. *A concordance to Beowulf*. Ithaca, N.Y.: Cornell University Press.
- Bhattacharjee, Y. 2004. DATAPOINT. *Science* 306(5702): 1679.
- Bickel, R., S. Weaver, T. Williams, and L. Lange. 1997. Opportunity, community, and teen pregnancy in an Appalachian state. *Journal of Educational Research* 90:175–81.
- Bickman, L., and D. J. Rog, eds. 1998. *Handbook of applied social research methods*. Thousand Oaks, Calif.: Sage.
- Bieri, D., R. A. Reeve, D. Champion, L. Addicoat, and J. B. Ziegler. 1990. The faces scale for

- the self-assessment of the severity of pain experienced by children: Development, initial validation, and preliminary investigation for ratio scale properties. *Pain* 41:139–50.
- Billiet, J., and G. Loosveldt. 1988. Improvement of the quality of responses to factual survey questions by interviewer training. *Public Opinion Quarterly* 52:190–211.
- Birdwell-Pheasant, D. 1984. Personal power careers and the development of domestic structure in a small community. *American Ethnologist* 11:699–717.
- Bizarro U. I. 1985. *Campesino: The diary of a Guatemalan Indian*. Translated and edited by J. D. Sexton. Tucson: University of Arizona Press.
- Bizarro U. I. 1992. *Ignacio: The diary of a Maya Indian of Guatemala*. Translated and edited by J. D. Sexton. Philadelphia: University of Pennsylvania Press.
- Bizarro U. I., and J. Sexton. 1981. *Son of Tecun Uman: A Maya Indian tells his life story*. Tucson: University of Arizona Press.
- Black, D., and A. J. Reiss, Jr. 1967. Patterns of behavior in police and citizen transactions. In *Studies in crime and law enforcement in major metropolitan areas. Field surveys III. U.S. President's Commission on Law Enforcement and Administration of Justice*. Vol. 2, pp. 1–39. Washington, D.C.: U.S. Government Printing Office.
- Blair, E. 1979. Interviewing in the presence of others. In *Improving interview method and questionnaire design: Response effects to threatening questions in survey research*, edited by N. M. Bradburn and S. Sudman, 134–46. San Francisco: Jossey-Bass.
- Blair, E., S. Sudman, N. M. Bradburn, and C. B. Stocking. 1977. How to ask questions about drinking and sex: Response effects in measuring consumer behavior. *Journal of Marketing Research* 14:316–21.
- Blalock, H. M. 1979. *Social statistics*. Rev. 2d ed. New York: McGraw-Hill.
- Boas, F. 1910. *Kwakiutl tales*. New York: Columbia University Press.
- Boas, F., ed. 1911. *Handbook of American Indian languages*. Washington, D.C.: Bureau of American Ethnology, Bulletin No. 40.
- Boas, F. 1917. *Folk-tales of Salishan and Sahaptin tribes*. Lancaster, Penn.: American Folk-Lore Society.
- Boas, F. 1928a. *Kerasan texts*. New York: The American Ethnological Society.
- Boas, F. 1928b. *Bella Bella texts*. New York: Columbia University Press.
- Boas, F. 1935–1943. *Kwakiutl tales*, New Series. New York: Columbia University Press.
- Boas, F. 1966 [1911]. *Introduction to handbook of American Indian languages*. Preston Holder, ed. Lincoln: University of Nebraska Press.
- Boas, F., and G. Hunt. 1902–1905. *Kwakiutl texts*. Leiden, The Netherlands: E. J. Brill.
- Boas, F., and G. Hunt. 1906. *Kwakiutl texts*, Second Series. *Memoirs of the American Museum of Natural History* 14. New York: G. E. Stechert.
- Bochner, S. 1971. The use of unobtrusive measures in cross-cultural attitudes research. In *A question of choice: An Australian Aboriginal dilemma*, edited by R. M. Berndt, 107–15. Netherlands, Western Australia: University of Western Australia Press.
- Bochner, S. 1972. An unobtrusive approach to the study of housing discrimination against Aborigines. *Australian Journal of Psychology* 24:335–37.
- Bochner, S. 1980. Unobtrusive observation in cross-cultural experimentation. In *Handbook of cross-cultural psychology*, edited by H. C. Triandis and J. W. Berry. Vol. 2, *Methodology*, 319–88. Boston: Allyn & Bacon.
- Bochner, S. 2000. Culture shock. In *Encyclopedia of psychology*. Vol. 2, edited by A. E. Kazdin, 410–13. Washington, D.C.: American Psychological Association.
- Bock, J. 2002. Evolutionary demography and intrahousehold time allocation: School attendance and child labor among the Okavango Delta peoples of Botswana. *American Journal of Human Biology* 14:206–21.

- Boeije, H. R. 2004. And then there were three: Why third persons are present in interviews and the impact on the data. *Field Methods* 16:3–32.
- Boelen, W.A.M. 1992. *Street corner society*. Cornerville revisited. *Journal of Contemporary Ethnography* 21(1): 11–51.
- Bogardus, E. S. 1933. A social distance scale. *Sociology and Social Research* 17 (January–February) 265: 271.
- Bogdan, R. 1972. *Participant observation in organizational settings*. Syracuse, N.Y.: Syracuse University Press.
- Bohannan, P. 1995. *How culture works*. New York: Free Press.
- Bolton, R. 1984. We all do it, but how? A survey of contemporary fieldnote procedure. In *Final report: Computers in ethnographic research* (Grant NIE–G–78–0062). Washington, D.C.: National Institute of Education (ERIC, No. ED 1. 310/2:248173).
- Bolton, R., and J. Vincke 1996. Risky sex and sexual cognition: The cartography of eros among Flemish gay men. *Journal of Quantitative Anthropology* 6:171–208.
- Booth, C., ed. 1902. *Life and labor of the people of London*. New York: Macmillan.
- Borchgrevink, A. 2003. Silencing language: Of anthropologists and interpreters. *Ethnography* 4:95–121.
- Borgatti, S. P. 1992a. *Anthropac 4.98*. Columbia, S.C.: Analytic Technologies.
- Borgatti, S. P. 1992b. *Anthropac 4.0 methods guide*. Columbia, S.C.: Analytic Technologies.
- Borgatti, S. P. 1993/1994. Cultural domain analysis. *Journal of Quantitative Anthropology* 4:261–78.
- Borgatti, S. P. 1999. Elicitation techniques for cultural domain analysis. In *Enhanced ethnographic methods*, edited by J. J. Schensul, M. D. LeCompte, B. K. Natasi, and S. P. Borgatti, 115–51. Walnut Creek, Calif.: AltaMira.
- Borgerhoff-Mulder, M. B., and T. M. Caro. 1985. The use of quantitative observational techniques in anthropology. *Current Anthropology* 26:323–36.
- Borofsky, R. 2005. *Yanomami: The fierce controversy and what we can learn from it*. Berkeley: University of California Press.
- Boruch, R. F., T. Victor, and J. S. Cecil. 2000. Resolving ethical and legal problems in randomized experiments. *Crime and Delinquency* 46:330–53.
- Boserup, E. 1970. *Women's role in economic development*. London: Allen & Unwin.
- Boster, J. S. 1985. Requiem for the omniscient informant: There's life in the old girl yet. In *Directions in cognitive anthropology*, edited by J. Dougherty, 177–98. Urbana: University of Illinois Press.
- Boster, J. S. 1986. Exchange of varieties and information between Aguaruna manioc cultivators. *American Anthropologist* 88:428–36.
- Boster, J. S. 1987. Agreement between biological classification systems is not dependent on cultural transmission. *American Anthropologist* 89:914–20.
- Boster, J. S., and J. C. Johnson. 1989. Form or function: A comparison of expert and novice judgments of similarity among fish. *American Anthropologist* 91:866–89.
- Boster, J. S., J. C. Johnson, and S. C. Weller. 1987. Social position and shared knowledge: Actors' perceptions of status, role, and social structure. *Social Networks* 9:375–87.
- Bourgois, P. I. 1990. Confronting anthropological ethics: Ethnographic lessons from Central America. *Journal of Peace Research* 27(1): 43–54.
- Bourgois, P. I. 1995. *In search of respect: Selling crack in El Barrio*. New York: Cambridge University Press.
- Boyle, E., Jr. 1970. Biological patterns in hypertension by race, sex, body weight, and skin color. *Journal of the American Medical Association* 213:1637–43.

- Bradburn, N. M. 1983. Response effects. In *Handbook of survey research*, edited by P. H. Rossi, J. D. Wright, and A. B. Anderson, 289–328. New York: Academic Press.
- Bradburn, N. M., S. E. Blair, and C. Stocking. 1978. Question threat and response bias. *Public Opinion Quarterly* 42:221–34.
- Bradburn, N. M., and S. Sudman, with the assistance of E. Blair et al. 1979. *Improving interview method and questionnaire design: Response effects to threatening questions in survey research*. San Francisco: Jossey-Bass.
- Bradley, C. 1997. Doing fieldwork with diabetes. *Cultural Anthropology Methods Journal* 9:1–7.
- Branstetter, K. B. 1977. A reconstruction of proto-Polynesian color terminology. *Anthropological Linguistics* 19(1): 1–25.
- Brettell, C. B. 1998. Fieldwork in the archives: Methods and sources in historical anthropology. In *Handbook of methods in cultural anthropology*, edited by H. R. Bernard, 513–46. Walnut Creek, Calif.: AltaMira.
- Brew, F. P., and D. R. Cairns. 2004. Do culture or situational constraints determine choice of direct or indirect styles in intercultural workplace conflicts? *International Journal of Intercultural Relations* 28:331–52.
- Brewer, D. D., and S. B. Garrett. 2001. Evaluation of interviewing techniques to enhance recall of sexual and drug injection partners. *Sexually Transmitted Diseases* 28:666–77.
- Brewer, D. D., S. B. Garrett, and G. Rinaldi. 2002. Free-listed items are effective cues for eliciting additional items in semantic domains. *Applied Cognitive Psychology* 16:343–58.
- Brewer, W. F. 1999. Schemata. In *The MIT encyclopedia of the cognitive sciences*, edited by R. A. Wilson and F. C. Keil, 729–30. Cambridge: MIT Press.
- Brewer, W. F. 2000. Bartlett's concept of the schema and its impact on theories of knowledge representation in contemporary cognitive psychology. In *Bartlett, culture and cognition*, edited by A. Saito, 69–89. Hove, England: Psychology Press.
- Bridges, F. S., and N. P. Coady. 1996. Affiliation, urban size, urgency, and cost of responses to lost letters. *Psychological Reports* 79:775–80.
- Bridgman, P. W. 1927. *The logic of modern physics*. New York: Macmillan. (Reprinted 1980, New York: Arno.)
- Brief, A. P., and L. Roberson. 1989. Job attitude organization: An exploratory study. *Journal of Applied Social Psychology* 19:717–27.
- Briere J., A. Downes, and J. Spensley. 1983. Summer in the city: Weather conditions and psychiatric emergency room visits. *Journal of Abnormal Psychology* 92:77–80.
- Brislin, R. W. 1970. Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology* 1(3): 185–216.
- Brislin, R. W., W. J. Lonner, and R. M. Thorndike. 1973. *Cross-cultural research methods*. New York: Wiley.
- Britt, C., G. Kleck, and D. J. Bordua. 1996. A reassessment of the D.C. gun law: Some cautionary notes on the use of interrupted time series designs for policy impact assessment. *Law and Society Review* 30:361–80.
- Broadbent, N. 1995. Accident claims lives of researchers in Russian Far East. *Anthropology Newsletter*, November, 39–40.
- Browne, K. E. 2001. Female entrepreneurship in the Caribbean: A multisite, pilot investigation of gender and work. *Human Organization* 60:326–42.
- Brugge, D. M. 1999. *The Navajo-Hopi land dispute: An American tragedy*. Albuquerque: University of New Mexico Press.
- Bryant, C. 1985. *Positivism in social theory and research*. New York: St. Martin's.
- Brymer, R. A. 1998. Hanging out with the good 'ole boys, gangsters, and other disreputable char-

- acters: Field research, quantitative research, and exceptional events. In *Doing ethnographic research: Fieldwork settings*, edited by S. Grills, 143–61. Thousand Oaks, Calif.: Sage.
- Bulmer, M. 1984. *The Chicago School of Sociology: Institutionalization, diversity, and the rise of sociological research*. Chicago: University of Chicago Press.
- Bulmer, M., ed. 1991. *Social research ethics: An examination of the merits of covert participant observation*. New York: Holmes and Meier.
- Bunzl, M. 1996. Franz Boas and the Humboldtian tradition. In *Volksgeist as method and ethic*, edited by G. W. Stocking, Jr., 17–78. Madison: University of Wisconsin Press.
- Burawoy, M. 1991. *Ethnography unbound: Power and resistance in the modern metropolis*. Berkeley: University of California Press.
- Burling, R. 1964. Cognition and componential analysis: God's truth or hocus-pocus? *American Anthropologist* 66:20–28.
- Burling, R. 2000 [1984]. *Learning a field language*. Prospect Heights, Ill.: Waveland.
- Burton, D. M. 1981. Automated concordances and word indexes: The process, the programs, and the products. *Computers and the Humanities* 15:139–54.
- Burton, M. L. 1968. Multidimensional scaling of role terms. Ph.D. diss., Stanford University.
- Burton, M. L. 1972. Semantic dimensions of occupation names. In *Multidimensional scaling: Applications in the behavioral sciences*, edited by A. K. Romney, R. N. Shepard, and S. B. Nerlove. Vol. 2, *Applications*, 55–72. New York: Seminar Press.
- Burton, M. L., and S. B. Nerlove. 1976. Balanced design for triad tests. *Social Science Research* 5:247–67.
- Burton, M. L., and A. K. Romney. 1975. A multidimensional representation of role terms. *American Ethnologist* 2(3): 397–407.
- Bush, A. J., and J. F. Hair, Jr. 1985. An assessment of the mall intercept as a data collection method. *Journal of Marketing Research* 22:158–67.
- Butler, B., and D. M. Turner. 1987. *Children and anthropological fieldwork*. New York: Plenum.
- Butler-Smith P., S. Cameron, and A. Collins. 1998. Gender differences in mate search effort: An exploratory economic analysis of personal advertisements. *Applied Economics* 30(10): 1277–85.
- Cabassa, L. J. 2003. Measuring acculturation: Where we are and where we need to go. *Hispanic Journal of Behavioral Sciences* 25:127–46.
- Cahn, P. S. 2002. Being the third generation in Tzintzuntzan. In *Chronicling cultures: Long-term field research in anthropology*, edited by R. V. Kemper and A. P. Royce, 313–28. Walnut Creek, Calif.: AltaMira.
- Cahnman, W. J. 1948. A note on marriage announcements in the *New York Times*. *American Sociological Review* 13:96–97.
- Cameron, S., and A. Collins. 1998. Sex differences in stipulated preferences in personal advertisements. *Psychological Reports* 82:119–23.
- Campbell, D. T. 1957. Factors relevant to the validity of experiments in social settings. *Psychological Bulletin* 54:297–312.
- Campbell, D. T. 1974. Evolutionary epistemology. In *The library of living philosophers*. Vol. 14, *The philosophy of Karl Popper*, Book 1, edited by P. A. Schlipp, 413–63. La Salle, Ill.: Open Court Publishing.
- Campbell, D. T. 1975. Degrees of freedom and the case study. *Comparative Political Studies* 8:178–93.
- Campbell, D. T. 1979. Degrees of freedom and the case study. In *Qualitative and quantitative methods in evaluation research*, edited by T. D. Cook and C. S. Reichart, 49–67. Beverly Hills, Calif.: Sage.
- Campbell, D. T., and R. F. Boruch. 1975. Making the case for randomized assignment to treat-

- ments by considering the alternatives: Six ways in which quasi-experimental evaluations in compensatory education tend to underestimate effects. In *Evaluation and experiment: Some critical issues in assessing social programs*, edited by C. A. Bennett and A. A. Lumsdaine, 195–296. New York: Academic Press.
- Campbell, D. T., and H. L. Ross. 1968. The Connecticut crackdown on speeding: Time-series data in quasi-experimental analysis. *Law and Society Review* 3:33–53.
- Campbell, D. T., and J. C. Stanley. 1966. *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
- Cannell, C. F., G. Fisher, and T. Bakker. 1961. Reporting of hospitalization in the Health Interview Survey. In *Health statistics*. Series D, No. 4. USDHEW, PHS. Washington, D.C.: U.S. Government Printing Office.
- Cannell, C. F., and F. J. Fowler. 1965. Comparison of hospitalization reporting in three survey procedures. In *Vital and health statistics*. Series 2, No. 8. Washington, D.C.: U.S. Government Printing Office.
- Cannell, C. F., and R. L. Kahn. 1968. Interviewing. In *The handbook of social psychology*, 2d ed., edited by G. Lindzey and E. Aronson. Vol. 2, *Research methods*, 526–95. Reading, Mass.: Addison-Wesley.
- Cannell, C. F., L. Oksenberg, and J. M. Converse. 1979. *Experiments in interview techniques. Field experiments in health reporting 1971–1977*. Ann Arbor: Institute for Social Research, University of Michigan.
- Cannell C. F., L. Oksenberg, G. Kalton, K. Bischooping, and F. J. Fowler. 1989. *New techniques for pretesting survey questions*. Final Report, Grant No. HS 05616. National Center for Health Services Research and Health Care Technology Assessment. Ann Arbor: Survey Research Center, University of Michigan.
- Cantril, H. 1965. *The pattern of human concerns*. New Brunswick, N.J.: Rutgers University Press.
- Cantwell, A.-M., E. Friedlander, and M. L. Tramm, eds. 2000. *Ethics and anthropology: Facing future issues in human biology, globalism, and cultural property*. New York: New York Academy of Sciences.
- Caplan, P. J., ed. 2003. *The ethics of anthropology: Debates and dilemmas*. New York: Routledge.
- Caplan, P. J., M. Crawford, J. S. Hyde, and J.T.E. Richeardson. 1997. *Gender differences in human cognition*. London: Oxford University Press.
- Carey, J. W., M. Morgan, and M. J. Oxtoby. 1996. Intercoder agreement in analysis of responses to open-ended interview questions: Examples from tuberculosis research. *Cultural Anthropology Methods Journal* 8:1–5.
- Carneiro, R. L. 1962. Scale analysis as an instrument for the study of cultural evolution. *Southwestern Journal of Anthropology* 18:149–69.
- Carneiro, R. L. 1970. Scale analysis, evolutionary sequences, and the rating of cultures. In *A handbook of method in cultural anthropology*, edited by R. Naroll and R. Cohen, 834–71. Garden City, N.Y.: Natural History Press.
- Carver, R. P. 1978. The case against statistical significance testing. *Harvard Educational Review* 48:378–99.
- Carver, R. P. 1993. The case against statistical significance testing, revisited. *Journal of Experimental Education* 61:287–92.
- Caserta, M. S., D. A. Lund, and M. F. Diamond. 1985. Assessing interviewer effects in a longitudinal study of bereaved elderly adults. *Journal of Gerontology* 40:637–40.
- Caspi, A., T. E. Moffitt, A. Thornton, D. Freedman, J. W. Amell, H. Harrington, J. Smeijers, and P. A. Silva. 1996. The life history calendar: Research and clinical assessment method for collecting retrospective event-history data. *International Journal of Methods in Psychiatric Research* 6:101–14.

- Cassell, J. 1982. Harm, benefits, wrongs, and rights in fieldwork. In *The ethics of social research: fieldwork, regulation, and publication*, edited by J. E. Sieber, 7–31. New York: Springer-Verlag.
- Cassell, J., ed. 1987. *Children in the field*. Philadelphia: Temple University Press.
- Casson, R. 1983. Schemata in cultural anthropology. *Annual Review of Anthropology* 12:429–62.
- Casterline, J. B., and V. C. Chidambaram 1984. The presence of others during the interview and the reporting of contraceptive knowledge and use. In *Survey analysis for the guidance of family planning programs*, edited by J. A. Ross and R. McNamara, 267–98. Liege, Belgium: Ordina Editions.
- Catania, J. A., D. Binson, J. Canchola, L. M. Pollack, W. Hauck, and T. J. Coates. 1996. Effects of interviewer gender, interviewer choice, and item wording on responses to questions concerning sexual behavior. *Public Opinion Quarterly* 60:345–75.
- Caulkins, D. D. 2001. Consensus, clines, and edges in Celtic cultures. *Cross-Cultural Research* 35:109–26.
- Chadsey-Rusch, J., and P. Gonzalez. 1988. Social ecology of the workplace: Employers' perceptions versus direct observations. *Research in Developmental Disabilities* 9:229–45.
- Chagnon, N. 1983. *Yanomamo. The fierce people*. 3rd ed. New York: Holt, Rinehart & Winston.
- Chalker, J., S. Ratanawijitrasin, N.T.K. Chuc, M. Petzold, and G. Tomson. 2004. Effectiveness of a multi-component intervention on dispensing practices at private pharmacies in Vietnam and Thailand—A randomized controlled trial. *Social Science and Medicine* 60:131–41.
- Chambers, R. 1991. Rapid (or relaxed) and participatory rural appraisal—notes on practical approaches and methods. *Qualitative Research Methods Newsletter* (Tata Institute of Social Research, Deonar, Bombay, India) (2), August:11–16.
- Chan, A., R. Madsen, and J. Unger. 1984. *Chen Village. The recent history of a peasant community in Mao's China*. Berkeley: University of California Press.
- Chapkis, W. 1997. *Live sex acts: Women performing erotic labor*. New York: Routledge.
- Chapman, P., R. B. Toma, R. V. Tuveson, and M. Jacob. 1997. Nutrition knowledge among adolescent high school female athletes. *Adolescence* 32:437–46.
- Chapple, E. D. 1940. Measuring human relations: An introduction to the study of the interaction of individuals. *Genetic Psychology Monographs* 22(1): 3–47.
- Chapple, E. D., and G. Donald, Jr. 1947. An evaluation of department store salespeople by the interaction chronograph. *Journal of Marketing* 12:173–85.
- Chick, G. 1994. Experience sampling in anthropological research. *Cultural Anthropology Methods Journal* 6(2): 4–6.
- Chisholm, J. S. 1983. *Navajo infancy: An ethological study of child development*. Hawthorne, N.Y.: Aldine.
- Choi, N., and D. R. Fuqua. 2003. The structure of the Bem Sex Role Inventory: A summary report of 23 validation studies. *Educational and Psychological Measurement* 63:872–87.
- Chomsky, N. 1957. *Syntactic structures*. 's-Gravenhage, The Netherlands: Mouton.
- Chomsky, N. 1969. *Deep structure, surface structure, and semantic interpretation*. Bloomington: Indiana University Linguistics Club.
- Chomsky, N. 1972. *Language and mind*. New York: Harcourt, Brace, Jovanovich.
- Chomsky, N. 1977. Recent contributions to the theory of innate ideas: Summary of oral presentation. In *Philosophy of Language*, edited by J. Searle, 121–29. Oxford: University Press.
- Chou, K.-L. 2000. Assessing Chinese adolescents' social support: The Multidimensional Scale of Perceived Social Support. *Personality and Individual Differences* 28(2): 299–307.
- Church, A. H. 1993. Estimating the effect of incentives on mail survey response rates: A meta-analysis. *Public Opinion Quarterly* 57(1): 62–79.
- Cialdini, R. B., R. J. Borden, A. Thorne, M. R. Walker, S. Freeman, and L. R. Sloan. 1976. Bask-

- ing in reflected glory: Three (football) field studies. *Journal of Personality and Social Psychology* 34:366–75.
- Claiborne, W. 1984. Dowry killings show social stress in India. *Washington Post*, 22 September, p. A1.
- Clark, S. J., and R. A. Desharnais. 1998. Honest answers to embarrassing questions: Detecting cheating in the randomized response model. *Psychological Methods* 3:60–168.
- Cliggett, L. 2002. Multigenerations and multidisciplines: Inheriting fifty years of Gwembe Tong research. In *Chronicling cultures: Long-term field research in anthropology*, edited by R. V. Kemper and A. P. Royce, 239–51. Walnut Creek, Calif.: AltaMira.
- Clignet, R. 1966. Urbanization and family structure in the Ivory Coast. *Comparative Studies in Society and History* 8:385–401.
- Clignet, R. 1970. *Many wives, many powers: Authority and power in polygynous families*. Evanston, Ill.: Northwestern University Press.
- Cochran, W. G. 1977. *Sampling techniques*. 2d ed. New York: Wiley.
- Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20:37–48.
- Cohen, J. 1988. *Statistical power analysis for the behavioral sciences*. 2d ed. Hillsdale, N.J.: Lawrence Erlbaum.
- Cohen, J. 1992. A power primer. *Psychological Bulletin* 112:155–59.
- Cohen, J. 1994. The earth is round ($p < .05$). *American Psychologist* 49:997–1003.
- Cohen-Kettenis, P. T., and L.J.G. Gooren. 1999. Transsexualism. A review of etiology, diagnosis and treatment. *Journal of Psychosomatic Research* 46:315–33.
- Cohn, E. G. 1990. Weather and crime. *British Journal of Criminology* 30:51–64.
- Cohn, E. G., and J. Rotton. 1997. Assault as a function of time and temperature: A moderator-variable time-series analysis. *Journal of Personality and Social Psychology* 72:1322–34.
- Colby, B. N. 1966. The analysis of culture content and the patterning of narrative concern in texts. *American Anthropologist* 68:374–88.
- Cole, D. 1983. The value of a person lies in his *herzenbildung*: Franz Boas' Baffin Island letter-diary, 1883–1884. In *Observers observed*, edited by G. W. Stocking, Jr., 13–52. Madison: University of Wisconsin Press.
- Coleman, J. S. 1966. *Equality of educational opportunity*. Washington, D.C.: U.S. Dept. of Health, Education, and Welfare, Office of Education/National Center for Education Statistics.
- Comrey, A. L. 1992. *A first course in factor analysis*. 2d ed. Hillsdale, N.J.: Lawrence Erlbaum.
- Comte, A. 1853. *The positive philosophy of Auguste Comte*. Edited and translated by H. Martineau. London: J. Chapman.
- Comte, A. 1975. *Auguste Comte and positivism: The essential writings*. Edited by G. Lenzer. New York: Harper & Row.
- Conklin, H. C. 1955. Hanunóo color categories. *Southwestern Journal of Anthropology* 11:339–44.
- Connolly, M. 1990. Adrift in the city: A comparative study of children in Bogota, Colombia, and Guatemala City. *Child and Youth Services* 14:129–49.
- Connolly, M., and J. Ennew, eds. 1996. Children out of place. Special Issue on Working and [sic] street children. *Childhood* 3, issue 2.
- Converse, J. M., and H. Schuman. 1974. *Conversations at random. Survey research as the interviewers see it*. New York: Wiley.
- Cook, S. W. 1975. A comment on the ethical issues involved in West, Gunn, and Chernicky's "Ubiquitous Watergate": An attributional analysis. *Journal of Personality and Social Psychology* 32:66–68.

- Cook, T. D., and D. T. Campbell. 1979. *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally College Publishing.
- Cooke, C. A. 2004. Young people's attitudes towards guns in America, Great Britain, and Western Australia. *Aggressive Behavior* 30:93–104.
- Coombs, C. H. 1964. *A theory of data*. New York: Wiley.
- Cornwall, A. 1992. Body mapping. *RRA Notes* 16. *Applications for Health*. 69–76.
- Corral-Verdugo, V. 1997. Dual "realities" of conservation behavior: Self-reports vs. observations of re-use and recycling behavior. *Journal of Environmental Psychology* 17(2): 135–45.
- Cottingham, J. 1999. *Descartes*. New York: Routledge.
- Couper, M. P., E. Singer, and R. Tourangeau. 2003. Understanding the effects of audio-CASI on self-reports of sensitive behavior. *Public Opinion Quarterly* 67:385–95.
- Cowan, G., and M. O'Brien. 1990. Gender and survival vs. death in slasher films—A content analysis. *Sex Roles* 23:187–96.
- Crabb, P. B., and D. Bielawski. 1994. The social representation of material culture and gender in children's books. *Sex Roles* 30:69–79.
- Csikszentmihalyi, M., and R. Larson. 1987. Validity and reliability of the experience-sampling method. Special issue: Mental disorders in their natural settings: The application of time allocation and experience-sampling techniques in psychiatry. *Journal of Nervous and Mental Disease* 175:526–36.
- Curtice, J., and N. Sparrow. 1997. How accurate are traditional quota opinion polls? *Journal of the Market Research Society* 39:433–48.
- Dadds, M. R., R. M. Rapee, and P. M. Barrett. 1994. Behavioral observation. In *International handbook of phobic and anxiety disorders in children and adolescents. Issues in clinical child psychology*, edited by T. H. Ollendick, N. King, and W. Yule, 349–64. New York: Plenum.
- Daley, W. R., A. Karpati, and M. Shelk. 2001. Needs assessment of the displaced population following the August 1999 earthquake in Turkey. *Disasters* 25:67–75.
- Daly, M., and M. Wilson. 1988. *Homicide*. New York: Aldine de Gruyter.
- D'Andrade, R. G. 1973. Cultural constructions of reality. In *Cultural illness and health*, edited by L. Nader and T. W. Maretzki, 115–27. Washington, D.C.: American Anthropological Association.
- D'Andrade, R. G. 1974. Memory and the assessment of behavior. In *Measurement in the social sciences*, edited by H. M. Blalock, Jr., 159–86. Chicago: Aldine.
- D'Andrade, R. G. 1991. The identification of schemas in naturalistic data. In *Person schemas and maladaptive interpersonal patterns*, edited by M. Horowitz, 279–301. Chicago: University of Chicago Press.
- D'Andrade, R. G. 1995. *The development of cognitive anthropology*. Cambridge: Cambridge University Press.
- D'Andrade, R. G., N. Quinn, S. B. Nerlove, and A. K. Romney. 1972. Categories of disease in American English and Mexican Spanish. In *Multidimensional scaling*, edited by A. K. Romney, R. Shepard, and S. B. Nerlove. Vol. 2, *Applications*, 9–54. New York: Seminar Press.
- D'Andrade, R. G., and C. Strauss, eds. 1992. *Human motives and cultural models*. New York: Cambridge University Press.
- Darwin, C. 1998 [1872]. *The expression of the emotions in man and animals*. 3rd. ed. New York: Oxford University Press.
- Davies, I., C. Davies, and G. Corbett. 1994. The basic colour terms of Ndebele. *African Languages and Cultures* 7(1): 36–48.
- Davis, D. 1986. Changing self-image: Studying menopausal women in a Newfoundland fishing

- village. In *Self, sex and gender in cross-cultural fieldwork*, edited by T. L. Whitehead and M. E. Conaway, 240–62. Urbana: University of Illinois Press.
- Davis, D. W. 1997. The direction of race of interviewer effects among African-Americans: Donning the black mask. *American Journal of Political Science* 41:309–23.
- Davis, J. A. 1971. *Elementary survey analysis*. Englewood Cliffs, N.J.: Prentice-Hall.
- Davis, K., and S. C. Weller. 1999. The effectiveness of condoms in reducing heterosexual transmission of HIV. *Family Planning Perspectives* 31(6): 272–79.
- Davis, N. Z. 1981. Printing and the people. In *Literacy and social development in the West*, edited by H. J. Graf, 69–95. Cambridge: Cambridge University Press.
- de Gheff, V. J. 1978. Hierarchical cluster analysis. In *Quantitative ethology*, edited by P. W. Colgan, 115–44. New York: Wiley.
- Dehavenon, A. L. 1978. Superordinate behavior in urban homes: A video analysis of request-compliance and food control behavior in two black and two white families living in New York City. Ph.D. diss., Columbia University.
- Dei, Kojo A. 2002. *Ties that bind. Youth and drugs in a black community*. Prospect Heights, Ill.: Waveland.
- de Leeuw, E., J. Hox, and S. Kef. 2003. Computer-assisted self-interviewing tailored for special populations and topics. *Field Methods* 15:223–51.
- de Leeuw, E., J. Hox, and G. Snijders. 1995. The effect of computer-assisted interviewing on data quality. A review. *Journal of the Market Research Society* 37:325–44.
- de Leeuw, E., and W. L. Nicholls, II. 1996. Technological innovations in data collection: Acceptance, data quality, and costs. *Sociological Research Online* 1(4). <http://www.socresonline.org.uk/socresonline/1/4/leeuw.html>.
- De Leon, G., J. A. Inciardi, and S. S. Martin. 1995. Residential drug abuse treatment research: Are conventional control designs appropriate for assessing treatment effectiveness? *Journal of Psychoactive Drugs* 27:85–91.
- Dellino, D. 1984. Tourism: Panacea or plight. Impacts on the quality of life on Exuma, Bahamas. Master's thesis, University of Florida.
- Deloria, V. 1969. *Custer died for your sins: An Indian manifesto*. New York: Macmillan.
- DeMaio, T. J. 1984. Social desirability and survey measurement: A review. In *Surveying subjective phenomena*. Vol. 2, edited by C. F. Turner and E. Martin, 257–82. New York: Russell Sage Foundation.
- DeMaio, T. J., and J. M. Rothgeb. 1996. Cognitive interviewing techniques: In the lab and in the field. In *Answering questions: Methodology for determining cognitive and communicative processes in survey research*, edited by N. Schwartz and S. S. Sudman, 177–96. San Francisco: Jossey-Bass.
- de Munck, V., N. Dudley, and J. Cardinale. 2002. Cultural models of gender in Sri Lanka and the United States. *Ethnology* 41:225–61.
- de Rada, V. D. 2001. Mail surveys using Dillman's TDM in a southern European country: Spain. *International Journal of Public Opinion Research* 13:159–72.
- Descartes, R. 1960 [1637]. *Discourse on method; and meditations*. New York: Liberal Arts Press.
- Descartes, R. 1993 [1641]. *Discourse on method and meditations on first philosophy*. Translated by Donald A. Cress. 3rd ed. Indianapolis: Hackett Publishing.
- Deutscher, I. 1973. *What we say, what we do*. Glenview, Ill.: Scott Foresman.
- DeVellis, R. F. 1991. *Scale development: Theory and applications*. Newbury Park, Calif.: Sage.
- DeWalt, B. R. 1979. *Modernization in a Mexican ejido*. New York: Cambridge University Press.
- DeWalt, K. M., and B. R. DeWalt. 2002. Participant observation: A guide for fieldworkers. Walnut Creek, Calif.: AltaMira.
- DeWalt, K. M., B. R. DeWalt, and C. B. Wayland. 1998. Participant observation. In *Handbook of*

- methods in cultural anthropology*, edited by H. R. Bernard, 259–99. Walnut Creek, Calif.: AltaMira.
- Dey, I. 1993. *Qualitative data analysis: A user-friendly guide for social scientists*. London: Routledge & Kegan Paul.
- Dickerson, S. S., M. A. Neary, and M. Hyche-Johnson. 2000. Native American graduate nursing students' learning experiences. *Journal of Nursing Scholarship* 32(2): 189–96.
- Dillman, D. A. 1978. *Mail and telephone surveys: The total design method*. New York: Wiley.
- Dillman, D. A. 1983. Mail and other self-administered questionnaires. In *Handbook of survey research*, edited by P. H. Rossi, J. D. Wright, and A. B. Anderson, 359–78. New York: Academic Press.
- Dilthey, W. 1985. *Introduction to the human sciences*. Princeton, N.J.: Princeton University Press. (Translation of the original from 1883.)
- Ditmar, H. 1991. Meanings of material possessions as reflections of identity: Gender and social-material position in society. *Journal of Social Behavior and Personality* 6:165–86.
- Divale, W. T. 1976. Female status and cultural evolution: A study in ethnographer bias. *Behavior Science Research* 11:169–212.
- Dixon, M., V. J. Roscigno, and R. Hodson. 2004. Unions, solidarity, and striking. *Social Forces* 83:3–33.
- Dobkin de Rios, M. 1981. Socioeconomic characteristics of an Amazon urban healer's clientele. *Social Science and Medicine* 15B:51–63.
- Dohrenwend, B. S., and S. A. Richardson. 1965. Directiveness and nondirectiveness in research interviewing: A reformulation of the problem. *Psychology Bulletin* 63:475–85.
- Doob, A. N., and A. E. Gross. 1968. Status of frustrator as an inhibitor of horn honking responses. *Journal of Social Psychology* 76:213–18.
- Dordick, G. A. 1996. More than refuge. *Journal of Contemporary Ethnography* 24:373–404.
- Doughty, P. 1979. A Latin American specialty in the world context: Urban primacy and cultural colonialism in Peru. *Urban Anthropology* 8:383–98.
- Douglas, M., ed. 1987. *Constructive drinking: Perspectives on drink from anthropology*. New York: Cambridge University Press.
- Dow, M. M., M. Burton, D. White, and K. Reitz. 1984. Galton's problem as network autocorrelation. *American Ethnologist* 11:754–70.
- Drake, S. 1978. *Galileo at work: His scientific biography*. Chicago: University of Chicago Press.
- Drass, K. 1980. The analysis of qualitative data: A computer program. *Urban Life* 9:332–53.
- Dressler, W. W. 1980. Ethnomedical beliefs and patient adherence to a treatment regimen: A St. Lucian example. *Human Organization* 39:88–91.
- Dressler, W. W., M. C. Balieiro, and J. E. dos Santos. 1997. The cultural construction of social support in Brazil: Associations with health outcomes. *Culture, Medicine and Psychiatry* 21:303–35.
- Dressler, W. W., M. C. Balieiro, and J. E. dos Santos. 2002. Cultural consonance and psychological distress. *Paidéia: Cadernos de Psicologia e Educação* 12:5–18.
- Dressler, W. W., R. P. Ribeiro, M. C. Balieiro, K. S. Oths, and J. E. dos Santos. 2004. Eating, drinking and being depressed: The social, cultural and psychological context of alcohol consumption and nutrition in a Brazilian community. *Social Science and Medicine* 59:709–20.
- Drury, C. C. 1990. Methods for direct observation of performance. In *Evaluation of human work: A practical ergonomics methodology*, edited by J. R. Wilson and E. N. Corlett, 35–57. New York: Taylor & Francis.
- Dugger, C. W. 2000. Kerosene, weapon of choice for attacks on wives in India. *New York Times*, 26 December:A1.

- Duncan, O. D. 1966. Path analysis: Sociological examples. *American Journal of Sociology* 72:1–16.
- Dunn-Rankin, P. 2004. *Scaling methods*. 2d ed. Mahwah, N.J.: Lawrence Erlbaum.
- Duranleau, D. 1999. Random sampling of regional populations: A field test. *Field Methods* 11:61–67.
- Durkheim, E. 1933 [1893]. *The division of labor in society*. Translated by G. Simpson. Glencoe, Ill.: The Free Press.
- Durkheim, E. 1951 [1897]. *Suicide. A study in sociology*. Glencoe, Ill.: Free Press.
- Durkheim, E. 1958. *Socialism and Saint-Simon*. Edited by A. Gouldner. Translated by C. Sattler. Yellow Springs, Ohio: Antioch Press.
- Dyk, W. 1938. *Son of Old Man Hat*. New York: Harcourt Brace.
- Dyk, W., and R. Dyk. 1980. *Left Handed; A Navaho autobiography*. New York: Columbia University Press.
- Dyl, J., and S. Wapner. 1996. Age and gender differences in the nature, meaning, and function of cherished possessions for children and adolescents. *Journal of Experimental Child Psychology* 62:340–77.
- Easlea, B. 1980. *Witch hunting, magic, and the new philosophy*. Atlantic Highlands, N.J.: Humanities Press.
- Eckman, P., ed. 1973. *Darwin and facial expression. A century of research in review*. New York: Academic Press.
- Eckman, P. 1980. *The face of man: Expression of universal emotions in a New Guinea village*. New York: Garland STPM Press.
- Egerton, R. B. 1966. Conceptions of psychosis in Four East African societies. *American Anthropologist* 68(2) Part 1: 408–25.
- Egerton, R. B. 1967. *The cloak of competence*. Berkeley and Los Angeles: University of California Press.
- Egerton, R. B., and S. Bercovici. 1976. The cloak of competence: Years later. *American Journal of Mental Deficiency* 80:485–97.
- Egerton, R. B., M. Bollinger, and B. Herr. 1984. The cloak of competence: After two decades. *American Journal of Mental Deficiency* 88:345–51.
- Egerton, R. B., and A. Cohen. 1994. Culture and schizophrenia: The DOSMD challenge. *British Journal of Psychiatry* 164:222–31.
- Egerton, R. B., and T. W. Ward. 1991. I gotta put my foot down. Richard T. Jarrett, III. In "I've seen it all." *Lives of older persons with mental retardation in the community*, edited by R. B. Egerton and M. A. Gaston, 123–49. Baltimore: Paul H. Brookes Publishing.
- Edmondson, B. 1988. This survey is garbage (analyzing trash to estimate populations). *American Demographics* 10:13–15.
- Edmonson, M. S., tr. and ed. 1986. *Heaven-born Mérida and its destiny: The Book of Chilam Balam of Chumayel*. Austin: University of Texas Press.
- Edwards, M., S. Thomsen, and C. Toroitich-Ruto. 2005. Thinking aloud to create better condom-use questions. *Field Methods* 17:183–99.
- Eibl-Eibesfeldt, I. 1989. *Human ethology*. New York: Aldine de Gruyter.
- Eisenstein, E. 1979. *The printing press as an agent of change: Communications and cultural transformations in early modern Europe*, 2 vols. Cambridge: Cambridge University Press.
- Ekman, P., E. R. Sorenson, and W. V. Friesen. 1969. Pan-cultural elements in facial displays of emotion. *Science* 164:86–88.
- El Guindi, F. 2004. *Visual anthropology*. Walnut Creek, Calif.: AltaMira.
- El Guindi, F., and A. Hernández Jiménez. 1986. *The myth of ritual. A native's ethnography of Zapotec life-crisis rituals*. Tucson: University of Arizona Press.

- Elliott, B. 2005. *Phenomenology and imagination in Husserl and Heidegger*. New York: Routledge.
- Ember, C. R., and M. Ember. 1992. Resource unpredictability, mistrust, and war: A cross-cultural study. *Journal of Conflict Resolution* 36:242–62.
- Ember, C. R., and M. Ember. 2001. *Cross-cultural research methods*. Walnut Creek, Calif.: AltaMira.
- Ember, C. R., and M. Ember. 2002. Father absence and male aggression. A re-examination of the comparative evidence. *Ethos* 29:296–314.
- Emerson, R. M., R. I. Fretz, and L. L. Shaw. 1995. *Writing ethnographic fieldnotes*. Chicago: University of Chicago Press.
- Ensminger, M. E., and K. Fothergill. 2003. A decade of measuring SES: What it tells us and where to go from here. In *Socioeconomic status, parenting, and child development*, edited by M. H. Bornstein and R. H. Bradley, 13–27. Mahwah, N.J.: Lawrence Erlbaum.
- Erasmus, C. J. 1955. Work patterns in a Mayo village. *American Anthropologist* 57:322–33.
- Erikson, K. T. 1967. A comment on disguised observation in sociology. *Social Problems* 14:366–73.
- Erikson, K. T. 1996. A response to Richard Leo. *The American Sociologist* 27:129–30.
- Evans-Pritchard, E. E. 1958 [1937]. *Witchcraft, oracles, and magic among the Azande*. Oxford: Oxford University Press.
- Evans-Pritchard, E. E. 1973. Some reminiscences and reflections on fieldwork. *Journal of the Anthropological Society of Oxford* 4:1–12.
- Fahim, H. M. 1982. *Indigenous anthropology in non-Western societies: Proceedings of a Burg-Wartenstein symposium*. Durham: University of North Carolina Press.
- Farmer, W. R. 1976. *The synoptic problem*. Dillsboro: Western North Carolina Press.
- Farnell, B. 1994. Ethno-graphics and the moving body. *Man* 29(4): 929–74.
- Farnell, B. 1995. *Do you see what I mean?: Plain Indian sign talk and the embodiment of action*. Austin: University of Texas Press.
- Farnell, B. 1996. Movement notation systems. In *The world's writing systems*, edited by P. T. Daniels and W. Bright, 1–25. New York: Oxford University Press.
- Farnell, B. 1999. Moving bodies, acting selves. *Annual Review of Anthropology* 28:341–73.
- Farnell, B., and L. R. Graham. 1998. Discourse-centered methods. In *Handbook of methods in cultural anthropology*, edited by H. R. Bernard, 411–57. Walnut Creek, Calif.: AltaMira.
- Farrington, J. M., M. G. Farrington, and D. Thomas. 1980. *A concordance and word-lists to the poems of Dylan Thomas*. Swansea, England: Ariel House.
- Feigl, H. 1980. Positivism. In *Encyclopaedia britannica*, Vol. 14. Chicago: Encyclopaedia Britannica, Inc.
- Feigl, H., and A. Blumberg. 1931. Logical positivism: A new movement in European philosophy. *Journal of Philosophy* 28:281–96.
- Feldman, K. D. 1994. Socioeconomic structures and mate selection among urban populations in developing regions. *Journal of Comparative Family Studies* 25(3): 329–43.
- Feldman, R. E. 1968. Response to compatriot and foreigner who seek assistance. *Journal of Personality and Social Psychology* 10:202–14.
- Fenno, R. 1990. *Watching politicians: Essays on participant observation*. Berkeley: Institute of Governmental Studies, University of California at Berkeley.
- Ferguson, R. B., ed. 1984. *Warfare, culture, and environment*. Orlando, Fla.: Academic Press.
- Festinger, L. A. 1957. *A theory of cognitive dissonance*. Stanford, Calif.: Stanford University Press.
- Fetterman, D. 1989. *Ethnography step by step*. Newbury Park, Calif.: Sage.
- Filip, J. C., M. E. Ming, R. M. Levy, O. J. Hoffstad, and D. J. Margolis. 2004. Mail surveys can

- achieve high response rates in a dermatology patient population. *Journal of Investigative Dermatology* 122:39–43.
- Fine, G. A., and K. L. Sandstrom. 1988. *Knowing children: Participant observation with minors*. Newbury Park, Calif.: Sage.
- Fink, J. C. 1983. CATI's first decade: The Chilton experience. *Sociological Methods and Research* 12:153–68.
- Finkel, S. E., T. M. Guterbock, and M. J. Borg. 1991. Race-of-interviewer effects in a preelection poll: Virginia 1989. *Public Opinion Quarterly* 55:313–30.
- Finkler, K. 1974. *Estudio comparativo de la economía de dos comunidades de México: El papel de la irrigación*. Mexico City: Instituto Nacional Indigenista.
- Fischer, C. 1982. *To dwell among friends: Personal networks in town and city*. Chicago: University of Chicago Press.
- Fisher, R. A. 1936. The use of multiple measurements in taxonomic problems. *Annals of Eugenics* 7:179–88.
- Fitzgerald, J., P. Gottschalk, and R. Moffitt. 1998. An analysis of the impact of sample attrition on the second generation of respondents in the Michigan panel study of income dynamics. *The Journal of Human Resources* 33:300–43.
- Fjellman, S. M., and H. Gladwin. 1985. Haitian family patterns of migration to South Florida. *Human Organization* 44:301–12.
- Flanagan, T. J., and D. R. Longmire, eds. 1996. *Americans view crime and justice: A national public opinion survey*. Newbury Park, Calif.: Sage.
- Fleisher, M. 1989. *Warehousing violence*. Newbury Park, Calif.: Sage.
- Fleisher, M. 1995. *Beggars and thieves: Lives of urban street criminals*. Madison: University of Wisconsin Press.
- Fleisher, M. 1998. *Dead end kids: Gang girls and the boys they know*. Madison: University of Wisconsin Press.
- Flores, E., S. L. Eyre, and S. G. Millstein. 1998. Sociocultural beliefs related to sex among Mexican American adolescents. *Hispanic Journal of Behavioral Sciences* 20(1): 60–82.
- Fluehr-Lobban, C. 1998. Ethics. In *Handbook of methods in cultural anthropology*, edited by H. R. Bernard, 173–202. Walnut Creek, Calif.: AltaMira.
- Fluehr-Lobban, C. 2002. *Ethics and the profession of anthropology dialogue for ethically conscious practice*. 2d ed. Walnut Creek, Calif.: AltaMira.
- Fortune, R. 1963 [1932]. *Sorcerers of Dobu. The social anthropology of the Dobu Islanders of the Western Pacific*. New York: E. P. Dutton & Co.
- Foster, G. M. 2002. A half-century of field research in Tzintzuntzan, Mexico: A personal view. In *Chronicling cultures: Long-term field research in anthropology*, edited by R. V. Kemper and A. P. Royce, 252–83. Walnut Creek, Calif.: AltaMira.
- Foster, G. M., T. Scudder, E. Colson, and R. V. Kemper, eds. 1979. *Long-term field research in social anthropology*. New York: Academic Press.
- Fowler, F. J. 1984. *Survey research methods*. Newbury Park, Calif.: Sage.
- Fox, R. J., M. R. Crask, and J. Kim. 1988. Mail survey response rate. *Public Opinion Quarterly* 52:467–91.
- Frake, C. O. 1961. The diagnosis of disease among the Subanum of Mindanao. *American Anthropologist* 63:113–32.
- Frake, C. O. 1962. The ethnographic study of cognitive systems. In *Anthropology and Human Behavior*, 72–85. Washington, D.C.: Anthropological Society of Washington.
- Frake, C. O. 1964. Notes on queries in anthropology. In *Transcultural studies in cognition*, edited by A. K. Romney and R. G. D'Andrade. *American Anthropologist* 66:Part II.
- Frank, S. H., K. C. Stange, D. Langa, and M. Workings. 1997. Direct observation of community-

- based ambulatory encounters involving medical students. *JAMA, The Journal of the American Medical Association* 278:712–16.
- Freedman, D., A. Thornton, D. Camburn, D. Alwin, and L. Young-DeMarco 1988. The life history calendar: A technique for collecting retrospective data. *Sociological Methodology* 18:37–68.
- Freedman, D. A. 2001. Ecological inference and the ecological fallacy. In *International encyclopedia of the social and behavioral sciences*, Vol. 6, edited by N. J. Smelser and P. B. Baltes, 4027–30. New York: Elsevier.
- Freeman, D. 1999. *The fateful hoaxing of Margaret Mead: A historical analysis of her Samoan research*. Boulder, Colo.: Westview.
- Freeman, L. C., A. K. Romney, and S. C. Freeman. 1987. Cognitive structure and informant accuracy. *American Anthropologist* 89:310–25.
- Frees, E. W. 2004. *Longitudinal and panel data: Analysis and applications in the social sciences*. New York: Cambridge University Press.
- Freilich, M., ed. 1977. *Marginal natives at work: Anthropologists in the field*. 2d ed. Cambridge, Mass.: Schenkman.
- Friedenberg, L. 1995. *Psychological testing. Design, analysis, and use*. Needham Heights, Mass.: Allyn & Bacon.
- Frisancho, A. R. 1990. *Anthropometric standards for the assessment of growth and nutritional status*. Ann Arbor: University of Michigan Press.
- Fry, D. P. 1990. Play aggression among Zapotec children: Implications for the practice hypothesis. *Aggressive Behavior* 16:321–40.
- Furbee, N. L., K. Maynard, J. J. Smith, R. A. Benfer, Jr., S. Quick, and L. Ross. 1996. The emergence of color cognition from color perception. *Journal of Linguistic Anthropology* 6:223–40.
- Furlow, C. 2003. Comparing indicators of knowledge within and between cultural domains. *Field Methods* 15:51–62.
- Furnham, A., and S. Bochner. 1986. *Culture shock: Psychological reactions to unfamiliar environments*. London: Methuen.
- Gaito, J. 1980. Measurement scales and statistics: Resurgence of an old misconception. *Psychological Bulletin* 87:564–67.
- Galilei, Galileo 1953 [1632]. *Dialogue concerning the two chief world systems, Ptolemaic & Copernican*. Berkeley: University of California Press.
- Galilei, Galileo. 1997 [1632]. *Galileo on the world systems: A new abridged translation and guide*. Translated by M. A. Finocchiaro. Berkeley: University of California Press.
- Gans, L. P., and C. S. Wood. 1985. Discriminant analysis as a method for differentiating potential acceptors of family planning: Western Samoa. *Human Organization* 44:228–33.
- Garro, L. C. 1986. Intracultural variation in folk medical knowledge: A comparison between curers and noncurers. *American Anthropologist* 88:351–70.
- Garro, L. C. 2000. Remembering what one knows and the construction of the past: A comparison of cultural consensus theory and cultural schema theory. *Ethos* 28:275–319.
- Gates, R., and P. Solomon. 1982. Research using the mall intercept: State of the art. *Journal of Advertising Research* 22:43–49.
- Gatewood, J. B. 1983a. Loose talk: Linguistic competence and recognition ability. *American Anthropologist* 85:378–86.
- Gatewood, J. B. 1983b. Deciding where to fish: The skipper's dilemma in southeast Alaska's salmon seining. *Coastal Zone Management Journal* 10:347–67.
- Gatewood, J. B. 1984. Familiarity, vocabulary size, and recognition ability in four semantic domains. *American Ethnologist* 11:507–27.

- Gatz, M., and M. Hurwicz. 1990. Are old people more depressed? Cross-sectional data on Center for Epidemiological Studies Depression Scale factors. *Psychology and Aging* 5:284–90.
- Gaulin, S. C., and J. S. Boster. 1990. Dowry as female competition. *American Anthropologist* 92:994–1005.
- Gearing, J. 1995. Fear and loving in the West Indies: Research from the heart. In *Taboo: Sex, identity, and erotic subjectivity in anthropological fieldwork*, edited by D. Kulick and M. Willson, 186–218. London: Routledge.
- Geertz, C. 1973. *The interpretation of cultures. Selected essays*. New York: Basic Books.
- Gerard, H. B., and G. C. Mathewson. 1966. The effects of severity of initiation on liking for a group: A replication. *Journal of Experimental Social Psychology* 2:278–87.
- Gigengack, R., ed. 2000. *Contemporary street ethnography*. Special section of *Focaal Tijdschrift voor Anthropologie* No. 36.
- Gilly, M. C. 1988. Sex roles in advertising: A comparison of television advertisements in Australia, Mexico, and the United States. *Journal of Marketing* 52:75–85.
- Gittelsohn, J., A. V. Shankar, K. P. West, and R. M. Ram. 1997. Estimating reactivity in direct observation studies of health behaviors. *Human Organization* 56:182–89.
- Gladwin, C. H. 1976. A view of plan puebla: An application of hierarchical decision models. *Journal of Agricultural Economics* 59:881–87.
- Gladwin, C. H. 1980. A theory of real life choice: Applications to agricultural decisions. In *Agricultural decision making*, edited by P. Barlett, 45–85. New York: Academic Press.
- Gladwin, C. H. 1983. Contributions of decision-tree methodology to a farming systems program. *Human Organization* 42:146–57.
- Gladwin, C. H. 1989. *Ethnographic decision tree modeling*. Newbury Park, Calif.: Sage.
- Gladwin, H. 1971. Decision making in the Cape Coast (Fante) fishing and fish marketing system. Ph.D. diss., Stanford University.
- Glaser, B. G., and A. Strauss. 1967. *The discovery of grounded theory: Strategies for qualitative research*. New York: Aldine.
- Glaser, J. M., and M. Gilens. 1997. Interregional migration and political resocialization: A study of racial attitudes under pressure. *Public Opinion Quarterly* 61:72–96.
- Glazer, M. 1975. Impersonal sex. In *Tearoom trade: Impersonal sex in public places*, edited by L. Humphreys, 213–22. Enl. ed. with a retrospect on ethical issues. Chicago: Aldine.
- Godoy, R. 2002. The life cycle, ecological, and economic determinants of spousal leisure sharing: Panel estimations from Tawahka Amerindians, Honduras. *Human Ecology* 30:317–37.
- Goffman, E. 1974. *Frame analysis*. New York: Harper & Row.
- Goldberg, D., S. T. Green, A. Taylor, M. Frischer, and N. McKeagancy. 1994. Comparison of four survey methods designed to estimate the prevalence of HIV among female prostitutes who inject drugs. *International Journal of STD and AIDS* 5:186–88.
- Golde, P., ed. 1986. *Women in the field: Anthropological experiences*. 2d ed. Berkeley: University of California Press.
- Goldsen, J. M. 1947. Analyzing the contents of mass communication: A step toward inter-group harmony. *International Journal of Opinion & Attitude Research* 1(4): 81–92.
- Goldstein, M. C. 1971. Stratification, polyandry, and family structure in central Tibet. *Southwestern Journal of Anthropology* 27:65–74.
- Gölge, Z. B., M. F. Yavuz, S. Mudderisoglou, and M. S. Yavuz. 2003. Turkish university students' attitudes toward rape. *Sex Roles* 49:653–61.
- González, N. S. 1986. The anthropologist as female head of household. In *Self, sex and gender in cross-cultural fieldwork*, edited by T. L. Whitehead and M. E. Conaway, 84–102. Urbana: University of Illinois Press.
- Good, K. (with D. Chanoff). 1991. *Into the heart*. New York: Simon & Schuster.

- Goode, E. 1996. Gender and courtship entitlement: Responses to personal ads. *Sex Roles* 34:141–69.
- Goode, W. J., and P. K. Hatt. 1952. *Methods in social research*. New York: McGraw-Hill.
- Goodenough, W. 1956. Componential analysis and the study of meaning. *Language* 32:195–216.
- Goodenough, W. 1965. Rethinking “status” and “role”: Toward a general model of the cultural organization of social relationships. In *The relevance of models for social anthropology. Association of social anthropology monographs I*, edited by M. Banton, 1–24. London: Tavistock.
- Goodman, L., and W. Kruskal. 1963. Measures of association for cross classifications III: Approximate sampling theory. *Journal of the American Statistical Association* 58:302–22.
- Goodwin, C. 1981. *Conversational organization: Interaction between speakers and hearers*. New York: Academic Press.
- Goodwin, C. 1986. Gesture as a resource for the organization of mutual orientation. *Semiotica* 62:29–49.
- Goodwin, C. 1994. Recording human interaction in natural settings. *Pragmatics* 3:181–209.
- Gorden, R. L. 1975. *Interviewing: Strategy, techniques, and tactics*. Homewood, Ill.: Dorsey.
- Gore-Felton, C., C. Koopman, and E. Bridges. 2002. An example of the maximizing survey return rates: Methodological issues for health professionals. *Evaluation and the Health Professions* 25:152–68.
- Gottschalk, L. A., and R. J. Bechtel. 1993. *Psychologic and neuropsychiatric assessment. Applying the Gottschalk-Gleser Content Analysis Method to verbal sample analysis using the Gottschalk-Bechtel Computer Scoring System*. Palo Alto, Calif.: Mind Garden.
- Gould, R. A., and P. B. Potter. 1984. Use-lives of automobiles in America: A preliminary archaeological view. In *Toward an ethnoarchaeology of modern America*, edited by R. A. Gould, 69–93. Providence, R.I.: Brown University, Department of Anthropology, Research Papers in Anthropology (No. 4).
- Gove, S., and G. H. Pelto. 1994. Focused ethnographic studies in the WHO programme for the control of acute respiratory infections. *Medical Anthropology* 15(4): 409–24.
- Graham, M. A. 2003. Adaptation of the weighed food record method to households in the Peruvian Andes and ethnographic insights on hunger. *Field Methods* 15:143–60.
- Graves, T. D., and C. A. Lave. 1972. Determinants of urban migrant Indian wages. *Human Organization* 31:47–61.
- Gravlee, C. C. 2002a. Mobile computer-assisted personal interviewing with handheld computers: The Entryware System 3.0. *Field Methods* 14:322–36.
- Gravlee, C. C. 2002b. Skin color, blood pressure, and the contextual effect of culture in southeastern Puerto Rico. Ph.D. diss., University of Florida.
- Gravlee, C. C., and William W. Dressler. 2005. Skin pigmentation, self-perceived color, and arterial blood pressure in Puerto Rico. *American Journal of Human Biology* 17(2): 195–206.
- Greene, P. D. 2001. Handheld computers as tools for writing and managing field data. *Field Methods* 13:181–97.
- Greenwald, A. G., R. González, R. J. Harris, and D. Guthrie. 1996. Effect sizes and p values: What should be reported and what should be replicated? *Psychophysiology* 33:175–83.
- Gregor, T. 1979. Short people. *Natural History* 88(2): 14–21.
- Gregson, S., T. Zhuwau, R. M. Anderson, and S. K. Chandiwana. 1998. Is there evidence for behaviour change in response to AIDS in rural Zimbabwe? *Social Science & Medicine* 46(3): 321–30.
- Gribble, J. N., H. G. Miller, and S. M. Rogers. 1999. Interview mode and measurement of sexual behaviors: Methodological issues. *Journal of Sex Research* 36:16–24.
- Grieco, E. M., and R. C. Cassidy. 2001. Overview of race and Hispanic origin. Census 2000 Brief

- C2KBR/01-1. Washington, D.C.: U.S. Bureau of the Census. <http://www.census.gov/prod/2001pubs/c2kbr01-1.pdf>.
- Griffin, J. H. 1961. *Black like me*. Boston: Houghton-Mifflin.
- Gross, D. R. 1984. Time allocation: A tool for the study of cultural behavior. *Annual Review of Anthropology* 13:519-58.
- Gross, D. R. 1992. *Discovering anthropology*. Mountain View, Calif.: Mayfield.
- Guest, G. 2000. Using Guttman scaling to rank wealth: Integrating quantitative and qualitative data. *Field Methods* 12:346-57.
- Guilmet, G. M. 1979. Instructor reaction to verbal and nonverbal-visual behavior in the urban classroom. *Anthropology and Education Quarterly* 10:254-66.
- Gujarati, D. N. 1995. *Basic econometrics*. 3rd ed. New York: McGraw-Hill.
- Gujarati, D. N. 2003. *Basic econometrics*. 4th ed. New York: McGraw-Hill.
- Gummerson, E. 2000. *Qualitative methods in management research*. 2d ed. Thousand Oaks, Calif.: Sage.
- Guttman, L. 1950. The basis for scalogram analysis. In *Studies in social psychology in World War II*, edited by S. A. Stouffer et al. Vol. 4, *Measurement and prediction*, 60-90. Princeton, N.J.: Princeton University Press.
- Hadaway, C. K., P. L. Marler, and M. Chaves. 1993. What the polls don't show: A closer look at U.S. church attendance. *American Sociological Review* 58:741-52.
- Hadaway, C. K., P. L. Marler, and M. Chaves. 1998. Overreporting church attendance in America: Evidence that demands the same verdict. *American Sociological Review* 63:122-30.
- Haddon, A. C., ed. 1901-1935. *Reports of the Cambridge Anthropological Expedition to Torres Straits*. Cambridge, England: The University Press.
- Halaby, C. N. 2004. Panel models in sociological research: Theory into practice. *Annual Review of Sociology* 30:507-44.
- Hall, E. T. 1963. A system of notation of proxemic behavior. *American Anthropologist* 65:1003-26.
- Hall, E. T. 1966. *The hidden dimension*. New York: Doubleday.
- Hamilton, W. L., J. T. Cook, W. W. Thompson, L. F. Buron, E. A. Frongillo, Jr., C. M. Olson, and C. A. Wehler. 1997a. *Household food security in the United States in 1995: Summary report of the food security measurement project*. Alexandria, Va.: Office of Analysis and Evaluation, Food and Consumer Service, U.S. Department of Agriculture.
- Hamilton, W. L., J. T. Cook, W. W. Thompson, L. F. Buron, E. A. Frongillo, Jr., C. M. Olson, and C. A. Wehler. 1997b. *Household food security in the United States in 1995: Technical report*. Alexandria, Va.: Office of Analysis and Evaluation, Food and Consumer Service, U.S. Department of Agriculture.
- Hammar, L. 2004. Bodies and methods in motion. *Practicing Anthropology* 26(Fall): 8-12.
- Hammel, E. A. 1962. Social rank and evolutionary position in a coastal Peruvian village. *Southwestern Journal of Anthropology* 18:199-215.
- Handlin, O. 1979. *Truth in history*. Cambridge, Mass.: Harvard University Press.
- Handwerker, W. P. 1989. *Women's power and social revolution: Fertility transition in the West Indies*. Newbury Park, Calif.: Sage.
- Handwerker, W. P. 1993. Simple random samples of regional populations. *Cultural Anthropology Methods Journal* 5(1): 12.
- Handwerker, W. P. 1996a. Constructing Likert scales: Testing the validity and reliability of single measures of multidimensional variables. *Cultural Anthropology Methods Journal* 8(1): 1-6.
- Handwerker, W. P. 1996b. Power and gender: Violence and affection experienced by children in Barbados, W. I. *Medical Anthropology* 17:101-28.

- Handwerker, W. P. 1997. Universal human rights and the problem of unbounded cultural meanings. *American Anthropologist* 99:799–809.
- Handwerker, W. P. 1998. Why violence? A test of hypotheses representing three discourses on the roots of domestic violence. *Human Organization* 57:200–208.
- Handwerker, W. P. 2001. *Quick ethnography: A guide to rapid multi-method research*. Walnut Creek, Calif.: AltaMira.
- Hanks, W. F. 1988. Grammar, style, and meaning in a Maya manuscript. Review of *Heaven born Mérida and its destiny: The book of Chilam Balam of Chumayel*. *International Journal of American Linguistics* 54:331–69.
- Hansen, A., and L. A. McSpadden. 1993. Self-anchoring scale, or ladder of life: A method used with diverse refugee populations. Unpublished manuscript.
- Hansley, W. E. 1974. Increasing response rates by choice of postage stamps. *Public Opinion Quarterly* 38:280–83.
- Harari, H., O. Harari, and R. V. White. 1985. The reaction to rape by American male bystanders. *Journal of Social Psychology* 125:653–58.
- Harburg, E., L. Gleibermann, P. Roeper, M. A. Schork, and W. J. Schull. 1978. Skin color, ethnicity and blood pressure I: Detroit blacks. *American Journal of Public Health* 68:1177–83.
- Harkness, J. A., F.J.R. van de Vijver, and P. P. Mohler, eds. 2003. *Cross-cultural survey methods*. Hoboken, N.J.: John Wiley.
- Harris, M. 1968. *The rise of anthropological theory*. New York: Thomas Crowell.
- Harris, M. 1980. *Cultural materialism: The struggles for a science of culture*. New York: Vintage Books.
- Harris, M., J. G. Consorte, J. Lang, and B. Byrne. 1993. Who are the Whites? Imposed census categories and the racial demography of Brazil. *Social Forces* 72:451–62.
- Harrison, A. 1993. Comparing nurses' and patients' pain evaluations: A study of hospitalized patients in Kuwait. *Social Science & Medicine* 36(5): 683–92.
- Harrison, E. 2000. Men, women and work in rural Zambia. *European Journal of Development Research* 12:53–71.
- Harrison, F. V., ed. 1997. *Decolonizing anthropology: Moving further toward an anthropology of liberation*. Arlington, Va.: Association of Black Anthropologists, American Anthropological Association.
- Harrison, G. G. 1976. Sociocultural correlates of food utilization and waste in a sample of urban households. Ph.D. diss., University of Arizona.
- Harshbarger, C. 1986. Agricultural credit in San Vito, Costa Rica. Master's thesis, University of Florida.
- Harshbarger, C. 1995. Farmer-herder conflict and state legitimacy in Cameroon. Ph.D. diss., University of Florida.
- Hartman, J. J. 1978. Social demographic characteristics of Wichita, Sedwick County. In *Metropolitan Wichita—Past, present, and future*, edited by G. Miller and J. Skaggs, 22–37. Lawrence: Kansas Regents Press.
- Hartman, J. J., and J. Hedblom. 1979. *Methods for the social sciences: A handbook for students and non-specialists*. Westport, Conn.: Greenwood.
- Hartmann, D. P., and D. D. Wood. 1990. Observational methods. In *International handbook of behavior modification therapy*, 2d ed., edited by A. S. Bellack, M. Hersen, and A. E. Kazdin, 107–38. New York: Plenum.
- Hartmann, P. 1994. Interviewing when the spouse is present. *International Journal of Public Opinion Research* 6:298–306.
- Harvey, S. M., and S. Thorburn Bird. 2004. What makes women feel powerful? An exploratory

- study of relationship power and sexual decision-making with African Americans at risk for HIV/STDs. *Women and Health* 39:1–18.
- Hatch, D., and M. Hatch. 1947. Criteria of social status as derived from marriage announcements in the *New York Times*. *American Sociological Review* 12:396–403.
- Hauser, R. M. 2005. Survey response in the long run: The Wisconsin Longitudinal Study. *Field Methods* 17:3–29.
- Hausman, D. B., and A. Hausman. 1997. *Descartes's legacy: Minds and meaning in early modern philosophy*. Toronto: University of Toronto Press.
- Hawkes, K., J. F. O'Connell, and N. G. Blurton-Jones. 1997. Hadza women's time allocation, offspring production, and the evolution of the long menopausal life span. *Current Anthropology* 38:551–57.
- Hawkes, K., J. F. O'Connell, and N. G. Blurton-Jones. 2001. Hunting and nuclear families. Some lessons from the Hadza about men's work. *Current Anthropology* 42:681–709.
- Haworth-Hoepfner, S. 2000. The critical shapes of body image: The role of culture and family in the production of eating disorders. *Journal of Marriage and the Family* 62:212–27.
- Hayek, F. A. von 1952. *The counter-revolution of science*. Glencoe, Ill.: Free Press.
- Hazan, A. R., H. L. Lipton, and S. A. Glantz. 1994. Popular films do not reflect current tobacco use. *American Journal of Public Health* 84:998–1000.
- Heath, D. B. 2000. *Drinking occasions: Comparative perspectives on alcohol and culture*. New York: Brunner/Mazel.
- Heath, S. B. 1972. *Telling tongues*. New York: Columbia University Press.
- Heberlein, T. A., and R. Baumgartner. 1978. Factors affecting response rates to mailed questionnaires: A quantitative analysis of the published literature. *American Sociological Review* 43:447–62.
- Heberlein, T. A., and R. Baumgartner. 1981. Is a questionnaire necessary in a second mailing? *Public Opinion Quarterly* 45:102–8.
- Heckathorn, D. D. 1997. Respondent-driven sampling: A new approach to the study of hidden populations. *Social Problems* 44:174–99.
- Heckathorn, D. D. 2002. Respondent-driven sampling II: Deriving valid population estimates from chain-referral samples of hidden populations. *Social Problems* 49:11–34.
- Heckathorn, D. D., and J. Jeffri. 2001. Finding the beat: Using respondent-driven sampling to study jazz musicians. *Poetics* 28:307–29.
- Hedge, A., and Y. H. Yousif. 1992. Effects of urban size, urgency, and cost on helpfulness: A cross-cultural comparison between the United Kingdom and the Sudan. *Journal of Cross-Cultural Psychology* 23(1): 107–15.
- Hedican, E. J. 1986. Sibling terminology and information theory: An hypothesis concerning the growth of folk taxonomy. *Ethnology* 25:229–39.
- Helm, J. 1980. Female infanticide, European diseases, and population levels among the Mackenzie Dene. *American Ethnologist* 7:259–85.
- Hemmings, A. B. 2004. *Coming of age in U.S. high schools: Economic, kinship, religious, and political crosscurrents*. Mahwah, N.J.: Lawrence Erlbaum.
- Henderson, J., and J. P. Harrington. 1914. *Ethnzoology of the Tewa Indians*. Smithsonian Institution, Bureau of American Ethnology, Bulletin 56. Washington, D.C.: Government Printing Office.
- Henley, N. M. 1969. A psychological study of the semantics of animal terms. *Journal of Verbal Learning and Verbal Behavior* 8:176–84.
- Henry, G. T. 1990. *Practical sampling*. Newbury Park, Calif.: Sage.
- Herd, G. 1992. "Coming out" as a rite of passage: A Chicago study. In *Gay culture in America. Essays from the field*, edited by G. Herdt, 29–68. Boston: Beacon Press.

- Herrera, C. D. 2001. Ethics, deception, and "those Milgram experiments." *Journal of Applied Philosophy* 18:245–56.
- Hertz, R. 1995. Separate but simultaneous interviewing of husbands and wives: Making sense of their stories. *Qualitative Inquiry* 4:429–51.
- Herzfeld, M. 1977. Ritual and textual structures: The advent of spring in rural Greece. In *Text and context*, edited by R. K. Jain, 29–45. Philadelphia: Institute for the Study of Human Issues.
- Herzog, A. R., and J. G. Bachman. 1981. Effects of questionnaire length on response quality. *Public Opinion Quarterly* 45:549–59.
- Herzog, S. 2004. The effect of motive on public perceptions of the seriousness of murder in Israel. *British Journal of Criminology* 44:771–82.
- Hewett, P. C., A. S. Erulkar, and B. S. Mensch. 2004. The feasibility of computer-assisted survey interviewing in Africa: Experience from two rural districts in Kenya. *Social Science Computer Review* 22:319–34.
- Hewitt, M. 2002. Attitudes toward interview mode and comparability of reporting sexual behavior by personal interview and audio computer-assisted self-interviewing: Analyses of the 1995 National Survey of Family Growth. *Sociological Methods and Research* 31:3–26.
- Hiatt, L. R. 1980. Polyandry in Sri Lanka: A test case for parental investment theory. *Man* 15:583–98.
- Hielscher, S., and J. Sommerfeld. 1985. Concepts of illness and the utilization of health-care services in a rural Malian village. *Social Science and Medicine* 21:469–81.
- Hirschman, E. C. 1987. People as products: Analysis of a complex marketing exchange. *Journal of Marketing* 51:98–108.
- Hochstim, J. R. 1967. A critical comparison of three strategies of collecting data from households. *Journal of the American Statistical Association* 62:976–89.
- Hodge, L. G., and D. Dufour. 1991. Cross-sectional growth of young Shipibo Indian children in eastern Peru. *American Journal of Physical Anthropology* 84:35–41.
- Hodges, H. A. 1952. *The philosophy of Wilhelm Dilthey*. London: Routledge and Paul.
- Hodson, R. 2004. A meta-analysis of workplace ethnographies. Race, gender, and employee attitudes and behaviors. *Journal of Contemporary Ethnography* 33:4–38.
- Holbrook, A. L., M. C. Green, and J. A. Krosnick. 2003. Telephone versus face-to-face interviewing of national probability samples with long questionnaires. Comparisons of respondent satisficing and social desirability response bias. *Public Opinion Quarterly* 67:79–125.
- Holland, D. 1985. From situation to impression: How Americans get to know themselves and one another. In *Directions in cognitive anthropology*, edited by J.W.D. Dougherty, 389–412. Urbana: University of Illinois Press.
- Holland, D., and N. Quinn. 1987. *Cultural models in language and thought*. Cambridge: Cambridge University Press.
- Holland, D., and D. Skinner. 1987. Prestige and intimacy: The cultural models behind Americans' talk about gender types. In *Cultural models in language and thought*, edited by D. Holland and N. Quinn, 78–111. New York: Cambridge University Press.
- Hollis, M. 1996. Philosophy of social science. In *The Blackwell companion to philosophy*, edited by N. Bunnin and E. P. Tsui-James, 358–87. Oxford: Blackwell.
- Honimichl, J. 1997. Real growth driven by overseas research. *Marketing News* 31: H2–H43.
- Horn, W. 1960. Reliability survey: A survey on the reliability of response to an interview survey. *Het PTT-Bedrijf (The Hague)* 10:105–56.
- Hornik, J. 1992. Tactile stimulation and consumer response. *Journal of Consumer Research* 19:449–58.

- Hornik, J., and S. Ellis. 1988. Strategies to secure compliance for a mall intercept interview. *Public Opinion Quarterly* 52:539–51.
- Horowitz, I. L. 1965. The life and death of Project Camelot. *Trans-Action* 3(1): 3–7b, 44–47.
- Houck, L. D. and L. C. Drickamer, eds. 1996. *Foundations of animal behavior: Classic papers with commentaries*. Chicago: University of Chicago Press.
- House, J. S., W. Gerber, and A. J. McMichael. 1977. Increasing mail questionnaire response: A controlled replication and extension. *Public Opinion Quarterly* 41:95–99.
- Houseman, M., and D. R. White 1998. Taking sides: Marriage networks and Dravidian kinship in Lowland South America. In *Transformations of kinship*, edited by M. Godelier, T. R. Trautmann, and F. E. Tjon Sie Fat, 214–43. Washington, D.C.: Smithsonian Institution Press.
- Howard, J. 1973 [1792]. *Prisons and lazarettos*. Vol. 1. *The state of the prisons in England and Wales: With preliminary observations, and an account of some foreign prisons and hospitals*. Montclair, N.J.: Patterson Smith.
- Howell, N. 1981. *Inferring infanticide from Hudson's Bay Company population data, 1829–1934*. Working Paper no. 26, Structural Analysis Programme, Department of Sociology, University of Toronto.
- Howell, N. 1990. *Surviving fieldwork*. Washington, D.C.: American Anthropological Association.
- HRAF. 1967. The HRAF quality-control sample universe. *Behavior Science Notes* 2:63–69.
- Hudelson, P. M. 1994. The management of acute respiratory infections in Honduras: A field test of the Focused Ethnographic Study (FES). *Medical Anthropology* 15:435–36.
- Hughes, W. W. 1984. The method to our madness: The garbage project methodology. In *Household refuse analysis: Theory, method, and applications in social science*, edited by W. L. Rathje and C. K. Rittenbaugh, 41–50. *American Behavioral Scientist* 28(1, entire issue).
- Hume, L., and J. Mulcock, eds. 2004. *Anthropologists in the field*. New York: Columbia University Press.
- Humphreys, L. 1975. *Tearoom trade: Impersonal sex in public places*. Enl. ed. with a retrospect on ethical issues. Chicago: Aldine.
- Hunt, M. M. 1997. *How science takes stock: The story of meta-analysis*. New York: Russell Sage Foundation.
- Hursh-César, G., and P. Roy, eds. 1976. *Third World surveys: Survey research in developing nations*. Delhi: Macmillan.
- Husserl, E. 1964 [1907]. *The idea of phenomenology*. Translated by W. P. Alston and G. Nakhnikian. The Hague: Nijhoff.
- Husserl, E. 1999. *The essential Husserl: Basic writings in transcendental phenomenology*. Edited by D. Welton. Bloomington: Indiana University Press.
- Hutchins, E. 1980. *Culture and inference*. Cambridge: Harvard University Press.
- Hutt, S. J., and C. Hutt. 1970. *Direct observation and measurement of behavior*. Springfield, Ill.: C. C. Thomas.
- Hyman, H. H., and W. J. Cobb. 1975. *Interviewing in social research*. Chicago: University of Chicago Press.
- Hymes, D. 1964. Discussion of Burling's paper. *American Anthropologist* 66:116–19.
- Hymes, D. 1976. Louis Simpson's "The deserted boy." *Poetics* 5:119–55.
- Hymes, D. 1977. Discovering oral performance and measured verse in American Indian narrative. *New Literary History* 8:431–57.
- Hymes, D. 1980a. Verse analysis of a Wasco text: Hiram Smith's "At'unaqa." *International Journal of American Linguistics* 46:65–77.
- Hymes, D. 1980b. Particle, pause, and pattern in American Indian narrative verse. *American Indian Culture and Research Journal* 4(4): 7–51.

- Hymes, D. 1981. *In vain I tried to tell you: Essays in Native American ethnopoetics*. Philadelphia: University of Pennsylvania Press.
- Hymes, V. 1987. Warm Springs Sahaptin narrative analysis. In *Native American discourse: Poetics and rhetoric*, edited by J. Sherzer and A. Woodbury, 62–102. Cambridge: Cambridge University Press.
- Ibeh, K.I.N., and J. K.-U. Brock. 2004. Conducting survey research among organisational populations in developing countries: Can the drop and collect technique make a difference? *International Journal of Marketing Research* 46:375–83.
- Ice, G. 2004. Technological advances in observational data collection: The advantages and limitations of computer-assisted data collection. *Field Methods* 16:352–75.
- ICMR and Dept. of Sociology, East Carolina University. 1993. *Final technical report for the coastal North Carolina socioeconomic study*. Submitted to the U.S. Dept. of the Interior, Minerals Management Service, by the Institute for Coastal and Marine Resources and East Carolina University, Department of Sociology, in cooperation with Impact Assessment, Inc., La Jolla, California.
- Ives, E. 1995. *The tape-recorded interview. A manual for fieldworkers in folklore and oral history*. 2d ed. Knoxville: University of Tennessee Press.
- Ivis, F. J., S. J. Bondy, and E. M. Adlaf. 1997. The effect of question structure on self-reports of heavy drinking: Closed-ended versus open-ended questions. *Journal of Studies on Alcohol* 58:622–24.
- Jackson, B. 1987. *Fieldwork*. Urbana: University of Illinois Press.
- Jacoby, H. 1995. The economics of polygyny in sub-Saharan Africa: Female productivity and the demand for wives in Côte d'Ivoire. *Journal of Political Economy* 103:938–71.
- Jaeger, R. M. 1984. *Sampling in education and the social sciences*. New York: Longmans, Green.
- James, J. E. 1997. Is habitual caffeine use a preventable cardiovascular risk factor? *The Lancet* 349:279–81.
- Jamieson, P., K. Hall Jamieson, and D. Romer. 2003. The responsible reporting of suicide in print journalism. *American Behavioral Scientist* 46:1643–60.
- Jaranson, J. M., J. Butcher, L. Halcón, D. R. Johnson, C. Robertson, K. Savik, M. Spring, and J. Westermeyer. 2004. Somali and Oromo refugees: Correlates of torture and trauma history. *American Journal of Public Health* 94:591–98.
- Jargowsky, P. A. 2005. The ecological fallacy. In *The encyclopedia of social measurement*, edited by K. Kempf-Leonard. Vol. 1, A–E. 715–22. Oxford: Elsevier/Academic Press.
- Jaskyte, K., and W. W. Dressler. 2004. Studying culture as an integral aggregate variable: Organizational culture and innovation in a group of nonprofit organizations. *Field Methods* 16:265–84.
- Jasso, G. 1998. Exploring the justice of punishments: Framing, expressiveness, and the just prison sentence. *Social Justice Research* 11:397–422.
- Javeline, D. 1999. Response effects in polite cultures. *Public Opinion Quarterly* 63:1–28.
- Jefferson, G. 1973. A case of precision timing in ordinary conversation: Overlapped tag-positioned address terms in closing sequences. *Semiotica* 9:47–96.
- Jenkins, C. 1981. Patterns of growth and malnutrition in Belize. *American Journal of Physical Anthropology* 56:169–78.
- Jenkins, J. H., and R. J., Barrett, eds. 2004. *Schizophrenia, culture, and subjectivity: The edge of experience*. New York: Cambridge University Press.
- Jensen, P. M., and R. B. Coombs. 1994. Health and behavioral predictors of success in an intensive smoking cessation program for women. *Women and Health* 21:57–72.
- Jobe, J. B., D. M. Keler, and A. F. Smith. 1996. Cognitive techniques in interviewing older people. In *Answering questions: Methodology for determining cognitive and communicative proc-*

- esses in survey research*, edited by N. Schwarz and S. Sudman, 197–219. San Francisco: Jossey-Bass.
- Johnson, A. 1975. Time allocation in a Machiguenga community. *Ethnology* 14:310–21.
- Johnson, A. 1978. *Quantification in anthropology*. Stanford, Calif.: Stanford University Press.
- Johnson, A. 1987. *The time allocation studies checklist*. Los Angeles: UCLA Time Allocation Project. (Also in *Cross-cultural studies of time allocation*. New Haven, Conn.: Human Relations Area Files.)
- Johnson, J. C. 1990. *Selecting ethnographic informants*. Newbury Park, Calif.: Sage.
- Johnson, J. C., and D. Griffith. 1995. Promoting sportfishing development in Puerto Rico: Travel agents' perceptions of the Caribbean. *Human Organization* 54:295–303.
- Johnson, R. K., P. Driscoll, and M. I. Goran. 1996. Comparison of multiple-pass 24-hour recall estimates of energy intake with total energy expenditure determined by the doubly labeled water method in young children. *Journal of the American Dietetic Association* 96:1140–44.
- Johnston, J., and C. Walton. 1995. Reducing response effects for sensitive questions: A computer-assisted self-interview with audio. *Social Science Computer Review* 13:304–19.
- Johnstone, B.K.F., K. Ferrara, and J. M. Bean. 1992. Gender, politeness, and discourse management in same-sex and cross-sex opinion-poll interviews. *Journal of Pragmatics* 18:405–30.
- Jones, D. J. 1973. The results of role-playing in anthropological research. *Anthropological Quarterly* 46:30–37.
- Jones, M., and M. A. Nies. 1996. The relationship of perceived benefits of and barriers to reported exercise in older African American women. *Public Health Nursing* 13:151–58.
- Joravsky, D. 1970. *The Lysenko affair*. Cambridge: Harvard University Press.
- Jordan, B. 1992a. *Birth in four cultures: A cross-cultural investigation of childbirth in Yucatan, Holland, Sweden, and the United States*. 4th exp. ed. Revised by R. Davis-Floyd. Prospect Heights, Ill.: Waveland.
- Jordan, B. 1992b. *Technology and social interaction: Notes on the achievement of authoritative knowledge in complex settings*. Technical Report #IRL 92–0027. Palo Alto, Calif.: Institute for Research on Learning.
- Jordan, B., and A. Henderson. 1993. *Interaction analysis: Foundations and practice*. Palo Alto, Calif.: Xerox Palo Alto Research Center and Institute for Research on Learning. Working paper.
- Jorgensen, D. 1989. *Participant observation*. Newbury Park, Calif.: Sage.
- Jussaume, R. A., Jr., and Y. Yamada. 1990. A comparison of the viability of mail surveys in Japan and the United States. *Public Opinion Quarterly* 54:219–28.
- Kadushin, C. 1968. Power, influence and social circles: A new methodology for studying opinion makers. *American Sociological Review* 33:685–99.
- Kahn, R. L., and C. F. Cannell. 1957. *The dynamics of interviewing*. New York: Wiley.
- Kahneman, D., A. B. Krueger, and D. A. Schkade. 2004. A survey method for characterizing daily life experience: The day reconstruction method. *Science* 306(5702): 1776–80.
- Kail, B. L., D. D. Watson, S. Ray, and National AIDS Research Consortium. 1995. Needle-using practices within the sex industry. *American Journal of Drug and Alcohol Abuse* 21:241–55.
- Kallan, J. E. 1998. Drug abuse-related mortality in the United States: Patterns and correlates. *American Journal of Drug and Alcohol Abuse* 24:103–17.
- Kane, E. W., and L. J. Macaulay. 1993. Interviewer gender and gender attitudes. *Public Opinion Quarterly* 57:1–28.
- Kane, S. 2001. AIDS and criminal justice. *Annual Review of Anthropology* 30:457–79.
- Kaplowitz, M. D., T. Hadlock, and R. Levinbe. 2004. A comparison of web and mail survey response rates. *Public Opinion Quarterly* 68:94–101.
- Kasl, S. V. 1996. The influence of the work environment on cardiovascular health: A historical,

- conceptual, and methodological perspective. *Journal of Occupational Health Psychology* 1:42–56.
- Katsurada, E., and Y. Sugihara. 1999. A preliminary validation of the Bem Sex Role Inventory in Japanese culture. *Journal of Cross-Cultural Psychology* 30(5): 641–45.
- Katz, D. 1942. Do interviewers bias polls? *Public Opinion Quarterly* 6:248–68.
- Katz, K., and C. Naré. 2002. Reproductive health knowledge and use of services among young adults in Dakar, Senegal. *Journal of Biosocial Science* 34:215–31.
- Kaufman, G., and V. C. Phua. 2003. Is ageism alive in date selection among men? Age requests among gay and straight men in Internet personal ads. *Journal of Men's Studies* 11:225–35.
- Kearney, M. H., S. Murphy, K. Irwin, and M. Rosenbaum. 1995. Salvaging self—A grounded theory of pregnancy on crack cocaine. *Nursing Research* 44:208–13.
- Kearney, R. 1996. *Paul Ricoeur: The hermeneutics of action*. Thousand Oaks, Calif.: Sage Publications.
- Keesing, R. M. 1978. *'Elota's story. The life and times of a Solomon Islands big man*. New York: St. Martin's.
- Keil, J. E., S. H. Sandifer, C. B. Loadholt, and E. Boyle, Jr. 1981. Skin color and education effects on blood pressure. *American Journal of Public Health* 71:532–34.
- Keil, J. E., H. A. Tyroler, S. H. Sandifer, and E. Boyle, Jr. 1977. Hypertension: Effects of social class and racial admixture. *American Journal of Public Health* 67:634–39.
- Keith, J., C. L. Fry, A. P. Glascock, C. Ikels, J. Dickerson-Putnam, H. Harpending, and P. Draper. 1994. *The aging experience: Diversity and commonality across cultures*. Thousand Oaks, Calif.: Sage.
- Kelloway, E. K. 1998. *Using LISREL for structural equation modeling: A researcher's guide*. Thousand Oaks, Calif.: Sage.
- Kelly, G. A. 1955. *The psychology of personal constructs*. New York: Norton.
- Kemper, R. V. 2002. From Student to Steward: Tzintzuntzan as extended community. In *Chronicling cultures: Long-term field research in anthropology*, edited by R. V. Kemper and A. P. Royce, 284–312. Walnut Creek, Calif.: AltaMira.
- Kemper, R. V., and A. P. Royce, eds. 2002. *Chronicling cultures: Long-term field research in anthropology*. Walnut Creek, Calif.: AltaMira.
- Kemph, B. T., and T. Kasser. 1996. Effects of sexual orientation of interviewer on expressed attitudes toward male homosexuality. *The Journal of Social Psychology* 136:401–3.
- Kempton, W. 1987. Two theories of home heat control. In *Cultural models in language and thought*, edited by D. Holland and N. Quinn, 222–42. New York: Cambridge University Press.
- Kempton, W., J. S. Boster, and J. A. Jartley. 1995. *Environmental values in American culture*. Cambridge: MIT Press.
- Kendall, C. 1983. Loose structure of family in Honduras. *Journal of Comparative Family Studies* 16:257–72.
- Kendall, C., E. Leontsini, E. Gil, F. Cruz, P. Hudelson, and P. Pelto. 1990. Exploratory ethnoontology. *Cultural Anthropology Methods Journal* 2:11.
- Kendon, A. 1981, ed. *Nonverbal communication, interaction, and gesture: Selections from semiotica*. The Hague: Mouton.
- Kenner, A. N., and G. Katsimaglis. 1993. Gender differences in proxemics: Taxi-seat choice. *Psychological Reports* 72:625–26.
- Kent, R. N., J. Kanowitz, K. D. O'Leary, and M. Cheiken. 1977. Observer reliability as a function of circumstances of assessment. *Journal of Applied Behavioral Analysis* 10:317–24.
- Kenyatta, J. 1938. *Facing Mt. Kenya*. London: Secker & Warburg.
- Kerlinger, F. N. 1973. *Foundations of behavioral research*. New York: Holt, Rinehart & Winston.

- Kessler, R. C., and E. Wethington. 1991. The reliability of life event reports in a community survey. *Psychological Medicine* 21:723–38.
- Ketterer, M. W., and G. H. Maercklein. 1991. Caffeinated beverage use among Type A male patients suspected of CAD/CHD: A mechanism for increased risk? *Stress Medicine* 7:119–24.
- Kiecker, P., and J. E. Nelson. 1996. Do interviewers follow telephone instructions? *Journal of the Market Research Society* 38:16–176.
- Kilbride, P. L. 1992. Unwanted children as a consequence of delocalization in modern Kenya. In *Anthropological research: Process and application*, edited by J. J. Poggie et al., 185–206. Albany: State University of New York Press.
- Killworth, P. D., and H. R. Bernard. 1974. CATIJ: A new sociometric technique and its application to a prison living unit. *Human Organization* 33:335–50.
- Killworth, P. D., H. R. Bernard, and C. McCarty. 1984. Measuring patterns of acquaintanceship. *Current Anthropology* 25:381–98.
- Killworth, P.R.P. 1997. Culture and power in the British army: Hierarchies, boundaries and construction. Ph.D. diss., University of Cambridge.
- Kim, A. I. 1985. Korean color terms: An aspect of semantic fields and related phenomena. *Anthropological Linguistics* 27(4): 425–36.
- Kimball, S. T., and W. T. Partridge. 1979. *The craft of community study: Fieldwork dialogues*. Gainesville: University of Florida Press.
- King, G. 1997. *A solution to the ecological inference problem: Reconstructing individual behavior from aggregate data*. Princeton, N.J.: Princeton University Press.
- King, N. 1994. The qualitative research interview. In *Qualitative methods in organizational research: A practical guide*, edited by C. Cassell and G. Symon, 14–36. Thousand Oaks, Calif.: Sage.
- Kinsey, A. C., W. B. Pomeroy, and C. E. Martin. 1948. *Sexual behavior in the human male*. Philadelphia: Saunders.
- Kirk, J., and M. Miller. 1986. *Reliability and validity in qualitative research*. Newbury Park, Calif.: Sage.
- Kirk, R. E. 1982. *Experimental design*. 2d ed. Monterey, Calif.: Brooks/Cole.
- Kiš, A. 2005. An analysis of the impact of AIDS on funeral culture in Malawi. Paper presented at the meetings of the Society for Applied Anthropology, Santa Fe, New Mexico.
- Klein, F. C. 1999. On sports: Academic dilemma. *The Wall Street Journal*, 2 April:W4.
- Kleinman, A. 1980. *Patients and healers in the context of culture: An exploration of the borderland between anthropology, medicine, and psychiatry*. Berkeley: University of California Press.
- Klonoff, E. A., and H. Landrine. 2000. Revising and improving the African American Acculturation Scale. *Journal of Black Psychology* 26(2): 235–61.
- Kluckhohn, K. 1945. The personal document in anthropological science. In *The use of personal documents in history, anthropology, and sociology*, edited by L. Gottschalk, C. Kluckhohn, and R. Angell, 79–176. New York: Social Science Research Council, Bulletin 53.
- Kneidinger, L. M., T. L. Maple, and S. A. Tross. 2001. Touching behavior in sport: Functional components, analysis of sex differences, and ethological considerations. *Journal of Nonverbal Behavior* 25:43–62.
- Knodel, J., N. Havanon, and A. Pramualratana. 1984. Fertility transition in Thailand: A qualitative analysis. *Population and Development Review* 10:297–315.
- Koivusaari, R. 2002. Horizontal and vertical interaction in children's computer-mediated communications. *Educational Psychology* 22:235–47.

- Koocher, G. P. 1977. Bathroom behavior and human dignity. *Journal of Personality and Social Psychology* 35:120–21.
- Koopmans, L. H. 1981. *An introduction to contemporary statistics*. Boston: Duxbury.
- Korn, J. H. 1997. *Illusions of reality: A history of deception in social psychology*. Albany: State University of New York Press.
- Kornblum, W. 1989. Introduction. In *In the field: Readings on the field research experience*, edited by C. D. Smith and W. Kornblum, 1–8. New York: Praeger.
- Korsching, P., J. Donnermeyer, and R. Burdge. 1980. Perception of property settlement payments and replacement housing among displaced persons. *Human Organization* 39:332–33.
- Kortendick, O. 1996. *Drei Schwestern un ihre Kinder. Rekonstruktion von Familiengeschichte und Identitätstransmission bei Indischen Nederlanders mit Hilfe computerunterstützter Inhaltsanalyse*. Canterbury, England: University of Kent, Center for Anthropology and Computing.
- Kraemer, H. C., and S. Thiemann. 1987. *How many subjects? Statistical power analysis in research*. Newbury Park, Calif.: Sage.
- Kreiss, L., and E. Stockton. 1980. Using the outline of cultural materials as a basis for indexing the content of ethnographic films. *Behavior Science Research* 15:281–93.
- Krieger, L. 1986. Negotiating gender role expectations in Cairo. In *Self, sex and gender in cross-cultural fieldwork*, edited by T. L. Whitehead and M. E. Conaway, 117–28. Urbana: University of Illinois Press.
- Krippendorff, K. 1980. *Content analysis: An introduction to its methodology*. Beverly Hills, Calif.: Sage.
- Kroeber, A. L. 1919. On the principle of order in civilization as exemplified by changes in women's fashions. *American Anthropologist* 21:235–63.
- Kronenfeld, D. B. 2004. Definitions of cross versus parallel: Implications for a new typology (an appreciation of A. Kimball Romney). In *Methods of comparison: Studies in honor of A. Kimball Romney*. Thousand Oaks, Calif.: Sage. Special Issue, *Cross-Cultural Research* 38(3): 249–69.
- Kronenfeld, D. B., J. Kronenfeld, and J. E. Kronenfeld. 1972. Toward a science of design for successful food service. *Institutions and Volume Feeding* 70(11): 38–44.
- Krosnick, J. A. 1999. Survey research. *Annual Review of Psychology* 50:537–67.
- Krueger, R. A. 1994. *Focus groups: A practical guide for applied research*. 2d ed. Thousand Oaks, Calif.: Sage.
- Krysan, M., and M. P. Couper. 2003. Race in the live and virtual interview: Racial deference, social desirability, and activation effects in attitude surveys. *Social Psychology Quarterly* 66:364–83.
- Kubey, R. W., R. Larson, and M. Csikszentmihalyi. 1996. Experience sampling method: Application to communication research questions. *Journal of Communications* 46:99–120.
- Kulick, D., and M. Willson, eds. 1995. *Taboo: Sex, identity, and erotic subjectivity in anthropological fieldwork*. London: Routledge.
- Kunin, T. 1955. The construction of a new type of attitude measure. *Personnel Psychology* 8:65–77.
- Kunitz, S. J., H. Temkin-Greener, D. Broudy, and M. Haffner. 1981. Determinants of hospital utilization and surgery on the Navajo Indian Reservation, 1972–1978. *Social Science and Medicine* 15B:71–79.
- Kunovich, R. S., and R. G. Rashid. 1992. Mirror training in three dimensions for dental students. *Perceptual and Motor Skills* 75:923–28.
- Kurasaki, K. S. 1997. Ethnic identity and its development among third-generation Japanese Americans. Ph.D. diss., Department of Psychology, DePaul University.

- Laban Moogi Gwako, E. 1997. The effects of women's land tenure security on agricultural productivity among the Maragoli of western Kenya. Ph.D. diss., Washington University.
- Labov, W. 1972. *Language in the inner city: Studies in the Black English vernacular*. Philadelphia: University of Pennsylvania Press.
- Labovitz, S. 1971a. The assignment of numbers to rank order categories. *American Sociological Review* 35:515–24.
- Labovitz, S. 1971b. The zone of rejection: Negative thoughts on statistical inference. *Pacific Sociological Review* 14:373–81.
- Labovitz, S. 1972. Statistical usage in sociology. *Sociological Methods and Research* 3:14–37.
- Ladd, E. C., and K. H. Bowman. 1997. *Public opinion about abortion*. Washington, D.C.: The American Enterprise Institute.
- La Flesche, F. 1963 [1900]. *The middle five: Indian schoolboys of the Omaha Tribe*. Madison: University of Wisconsin Press.
- Lakoff, G., and M. Johnson. 2003 [1980]. *Metaphors we live by*. Chicago: University of Chicago Press.
- Lam, T.C.M., K. Green, and C. Bordignon. 2002. Effects of item grouping and position of the “Don't Know” option on questionnaire response. *Field Methods* 14:418–32.
- Lambert, E., R. S. Ashery, and R. H. Needle, eds. 1995. *Qualitative methods in drug abuse and HIV research*. Rockville, Md.: National Institute on Drug Abuse, Division of Epidemiology and Prevention Research.
- Lame Deer, J., and R. Erdoes. 1972. *Lame Deer, seeker of visions*. New York: Simon & Schuster.
- Landauer, T. K., and J.W.M. Whiting. 1964. Infantile stimulation and adult stature of human males. *American Anthropologist* 66:1007–28.
- Lang, H., P. Challenor, and P. D. Killworth. 2004. A new addition to the family of space sampling methods. *Field Methods* 16:55–69.
- Lange, R., ed. 1975. *Laban's principles of dance and movement notation*. 2d ed. Boston: Plays.
- La Pierre, R. T. 1934. Attitudes versus actions. *Social Forces* 13:230–37.
- Lara-Cantú, M. A., and R. Navarro-Arias. 1987. Self-description of Mexican college students in response to the Bem Sex Role Inventory and other sex role items. *Journal of Cross-Cultural Psychology* 18(3): 331–44.
- Laraia, B. A., A. M. Siega-Riz, J. S. Kaufman, and S. J. Jones. 2004. Proximity of supermarkets is positively associated with diet quality index for pregnancy. *Preventive Medicine* 39:869–75.
- Lastrucci, C. L. 1963. *The scientific approach*. Cambridge: Schenkman.
- Latané, B., and J. M. Darley. 1968. Group inhibition of bystander intervention in emergencies. *Journal of Personality and Social Psychology* 10:215–21.
- Lauder, M. 2003. Covert participant observation of a deviant community: Justifying the use of deception. *Journal of Contemporary Religion* 18:185–96.
- Lauder, W., P. A. Scott, and A. Whyte. 2001. Nurses' judgements of self-neglect: A factorial survey. *International Journal of Nursing Studies* 38:601–8.
- Laurent, C., K. Seck, N. Coumba, T. Kane, N. Samb, A. Wade, F. Liégeois, S. Mboup, I. Ndoye, and E. Delaporte. 2003. Prevalence of HIV and other sexually transmitted infections, and risk behaviours in unregistered sex workers in Dakar, Senegal. *AIDS* 17(12): 1811–16.
- Lazarsfeld, P. F. 1954. *Mathematical thinking in the social sciences*. Glencoe, Ill.: Free Press.
- Lazarsfeld, P. F. 1982. *The varied sociology of Paul F. Lazarsfeld: Writing*. New York: Columbia University Press.
- Lazarsfeld, P. F., A. Pasanella, and M. Rosenberg, eds. 1972. *Continuities in the language of social research*. New York: The Free Press.
- Lazarsfeld, P. F., and M. Rosenberg. 1955. *The language of social research*. Glencoe, Ill.: Free Press.

- Lea, K. L. 1980. Francis Bacon. *Encyclopaedia britannica*, Vol. 2. Chicago: Encyclopaedia Britannica Inc.
- Leach, E. R. 1945. Jinglypaw kinship terminology: An experiment in ethnographic algebra. *Journal of the Royal Anthropological Institute of Great Britain and Ireland* 75:59–72.
- Leach, E. R. 1967. An anthropologist's reflection on a social survey. In *Anthropologists in the field*, edited by D. C. Jongmans and P. C. Gutkind, 75–88. Assen, The Netherlands: Van Gorcum.
- LeCompte, M. D. 1978. Learning to work: The hidden curriculum of the classroom. *Anthropology and Education Quarterly* 9:22–37.
- LeCompte, M. D., and J. Preissle (with R. Tesch). 1993. *Ethnography and qualitative design in educational research*. 2d ed. San Diego: Academic Press.
- Lees, S. H. 1986. Coping with bureaucracy: Survival strategies in irrigated agriculture. *American Anthropologist* 88:610–22.
- Lehman, F. K. 1992. Relationship between genealogical and terminological structure in kinship. *Journal of Quantitative Anthropology* 4:95–22.
- Lehner, P. N., ed. 1996. *Handbook of ethological methods*. 2d ed. New York: Cambridge University Press.
- Leith, L. M. 1988. Choking in sports: Are we our own worst enemies? *International Journal of Sport Psychology* 19:59–64.
- Lenski, G. E., and J. C. Leggett. 1960. Caste, class and deference in the research interview. *American Journal of Sociology* 65:463–67.
- Leslie, P. W., R. Dyson-Hudson, E. A. Lowoto, and J. Munyesi. 1999. Appendix 1. Ngisonyoka event calendar. In *Turkana herders of the dry savanna: Ecology and biobehavioral response of nomads to an uncertain environment*, edited by M. A. Little and P. W. Leslie, 375–78. New York: Oxford University Press.
- Levine, R. V. 1997. *A geography of time*. New York: Basic Books.
- Levine, R. V., and K. Bartlett. 1984. Pace of life, punctuality, and coronary heart disease in six countries. *Journal of Cross-Cultural Psychology* 15:233–55.
- Levinson, D., ed. 1978. *A guide to social theory: Worldwide cross-cultural tests*. New Haven, Conn.: HRAF Press.
- Levinson, D. 1989. *Family violence in cross-cultural perspective*. Newbury Park, Calif.: Sage.
- Levy, P. S., and S. Lemeshow. 1999. *Sampling of populations: Methods and applications*. 3rd ed. New York: Wiley.
- Levy-Storms, L. 1998. The influence of communication networks on the diffusion of mammography screening among older Pacific Islander women. Ph.D. diss., University of California, Los Angeles.
- Levy-Storms, L. n.d. Social network influences on mammography screening behaviors among older Samoan women in Los Angeles County. Unpublished manuscript.
- Levy-Storms, L., J. F. Schnelle, and S. F. Simmons. 2002. A comparison of methods to assess nursing home residents' unmet needs. *The Gerontologist* 42:454–61.
- Levy-Storms, L., and S. P. Wallace. 2003. Use of mammography screening among older Samoan women in Los Angeles County: A diffusion network approach. *Social Science and Medicine* 57:987–1000.
- Lewin, E., and W. Leap, eds. 1996. *Out in the field: Reflections of lesbian and gay anthropologists*. Urbana: University of Illinois Press.
- Lewis, O. 1961. *The children of Sánchez*. New York: Random House.
- Lewis, O. 1965. *La Vida: A Puerto Rican family in the culture of poverty—San Juan and New York*. New York: Random House.

- Lewis, R. B. 2004. NVivo 2.0 and Atlas/ti 5.0: A comparative review of two popular qualitative data-analysis programs. *Field Methods* 16:439–69.
- Lieberman, D., and W. W. Dressler. 1977. Bilingualism and cognition of St. Lucian disease terms. *Medical Anthropology* 1:81–110.
- Lieske, J. 1993. Regional subcultures of the United States. *The Journal of Politics* 55:888–913.
- Lightcap, J. L., J. A. Kurland, and R. L. Burgess. 1982. Child abuse: A test of some predictions from evolutionary theory. *Ethology and Sociobiology* 3:61–67.
- Likert, R. 1932. A technique for the measurement of attitudes. *Archives of Psychology*, Vol. 22, No. 140.
- Lincoln, Y. S., and E. G. Guba 1985. *Naturalistic inquiry*. Beverly Hills, Calif.: Sage.
- Liu, J. H., B. Bonzon-Liu, and M. Pierce-Guarino. 1997. Common fate between humans and animals? The dynamical systems theory of groups and environmental attitudes in the Florida Keys. *Environment and Behavior* 29:87–124.
- Livingstone, S., and P. Lunt. 1993. Savers and borrowers: Strategies of personal financial management. *Human Relations* 46:963–85.
- Locke, J. 1996 [1690]. *An essay concerning human understanding*. Abridged. Edited by K. P. Winkler. Indianapolis: Hackett Publishing.
- Loether, H. J., and D. G. McTavish. 1993. *Descriptive and inferential statistics*. Boston: Allyn & Bacon.
- Lofland, J. H. 1971. *Analyzing social settings. A guide to qualitative observation and analysis*. Belmont, Calif.: Wadsworth.
- Lofland, J. H. 1976. *Doing social life*. New York: Wiley.
- Lofland, L. H. 1983. Understanding urban life: The Chicago legacy. *Urban Life* 11:491–511.
- Loftin, C., D. McDowall, B. Wiersma, and T. J. Cottey. 1991. Effects of restrictive licensing of handguns on homicide and suicide in the District of Columbia. *New England Journal of Medicine* 325(23): 1615–20.
- Loftus, E. F., and W. Marburger. 1983. Since the eruption of Mt. St. Helens, has anyone beaten you up? Improving the accuracy of retrospective reports with landmark events. *Memory and Cognition* 11:114–20.
- Longabaugh, R. 1980. The systematic observation of behavior in naturalistic settings. In *Handbook of cross-cultural psychology*, edited by H. C. Triandis and J. W. Berry. Vol. 2, *Methodology*, 57–126. Boston: Allyn & Bacon.
- Loo, R., and K. Thorpe. 1998. Attitudes toward women's roles in society: A replication after 20 years. *Sex Roles* 39:903–12.
- Low, S. M., and D. Lawrence-Zúñiga 2003. *The anthropology of space and place: Locating culture*. Malden, Mass.: Blackwell.
- Lowe, J.W.G., and E. D. Lowe. 1982. Cultural pattern and process: A study of stylistic change in women's dress. *American Anthropologist* 84:521–44.
- Lowie, R. H. 1940. Native languages as ethnographic tools. *American Anthropologist* 42:81–89.
- Lowrey, T. M., and C. Otnes. 1994. Construction of a meaningful wedding: Differences in the priorities of brides and grooms. In *Gender issues and consumer behavior*, edited by J. A. Costa, 164–83. Thousand Oaks, Calif.: Sage.
- Lucretius Carus, Titus. 1998. *De rerum natura. (On the nature of the universe)*. Translated by R. Melville. New York: Oxford University Press.
- Lueptow, L. B., S. L. Moser, and B. F. Pendleton. 1990. Gender and response effects in telephone interviews about gender characteristics. *Sex Roles* 22:29–42.
- Lundberg, G. A. 1964. *Foundations of sociology*. New York: David McKay.
- Lurie, N. 1961. *Mountain Lone Wolf Woman, Sister of Crashing Thunder: The autobiography of a Winnebago woman*. Ann Arbor: University of Michigan Press.

- Lynd, R. S., and H. M. Lynd. 1929. *Middletown. A study in contemporary American culture*. New York: Harcourt, Brace & Co.
- MacClancy, J. 2002. *Exotic no more: Anthropology on the front lines*. Chicago: University of Chicago Press.
- Mach, E. 1976. *Knowledge and error: Sketches on the psychology of enquiry*. Edited by B. McGuiness. Translated by T. J. McCormack and P. Foulkes. Boston: D. Reidel.
- Madden, J. M., J. D. Quick, and D. Ross-Degnan. 1997. Undercover careseekers: Simulated clients in the study of health provider behavior in developing countries. *Social Science and Medicine* 45:1465–82.
- Magaña, J. R., M. Burton, and J. Ferreira-Pinto. 1995. Occupational cognition in three nations. *Journal of Quantitative Anthropology* 5:149–68.
- Mahaffy, K. A. 1996. Cognitive dissonance and its resolution: A study of lesbian Christians. *Journal for the Scientific Study of Religion* 35:392–402.
- Malinowski, B. 1961 [1922]. *Argonauts of the western Pacific*. New York: Dutton.
- Malinowski, B. 1967. *A diary in the strict sense of the term*. New York: Harcourt, Brace & World.
- Manzo, J. 1996. Taking turns and taking sides: Opening scene from two jury deliberations. *Social Psychology Quarterly* 59:107–25.
- Margolis, M. 1984. *Mothers and such*. Berkeley: University of California Press.
- Markie, P. J. 1986. *Descartes' gambit*. Ithaca, N.Y.: Cornell University Press.
- Markowitz, F., and M. Ashkenazi. 1999. *Sex, sexuality, and the anthropologist*. Urbana: University of Illinois Press.
- Marquis, G. S. 1990. Fecal contamination of shanty town toddlers in households with non-corralled poultry, Lima, Peru. *American Journal of Public Health* 80:146–50.
- Marriott, B. 1991. The use of social networks by naval officers' wives. Ph.D. diss., University of Florida.
- Marsh, V. M., W. M. Mutemi, A. Willetts, K. Bayah, S. Were, A. Ross, and K. Marsh. 2004. Improving malaria home treatment by training drug retailers in rural Kenya. *Tropical Medicine and International Health* 9:451–60.
- Martin, J. 1981. Relative deprivation: A theory of distributive injustice for an era of shrinking resources. In *Research in organizational behavior*, Vol. 3, edited by L. L. Cummings and B. M. Staw, 53–107. Greenwich, Conn.: JAI Press.
- Martin, J. L., and L. Dean. 1993. Developing a community sample of gay men for an epidemiological study of AIDS. In *Researching sensitive topics*, edited by C. M. Renzetti and R. M. Lee, 82–100. Newbury Park, Calif.: Sage.
- Matarazzo, J. 1964. Interviewer mm-humm and interviewee speech duration. *Psychotherapy: Theory, Research and Practice* 1:109–14.
- Mathews, H. F. 1985. The Weeping Woman: Variation and homogeneity in folk theories of gender in a Mexican community. Paper read at the annual meetings of the American Anthropological Association, Washington, D.C.
- Mathews, H. F. 1992. The directive force of morality tales in a Mexican community. In *Human motives and cultural models*, edited by R. G. D'Andrade and C. Strauss, 127–62. New York: Cambridge University Press.
- Mathews, J. J. 1968 [1932]. *Wah'Kon-Tah: The Osage and the white man's road*. Norman: University of Oklahoma Press.
- Matland, R. E., and D. T. Studlar. 1996. The contagion of women candidates in single-member district and proportional representation electoral systems: Canada and Norway. *Journal of Politics* 58:707–33.
- Max-Planck Institute. 2002. The human life table database 2002. <http://www.lifetable.de/>.

- McCall, G. 1978. *Observing the law: Field methods in the study of crime and the criminal justice system*. New York: The Free Press.
- McCarroll, J. E., A. S. Blank, and K. Hill. 1995. Working with traumatic material: Effects on Holocaust Memorial Museum staff. *American Journal of Orthopsychiatry* 65:66–75.
- McClendon, M. 1991. Acquiescence and recency response-order effects in interview surveys. *Sociological Methods and Research* 20:60–103.
- McDonald, D. D., and R. G. Bridge. 1991. Gender stereotyping and nursing care. *Research in Nursing and Health* 14:373–78.
- McDonald, L. 1993. *The early origins of the social sciences*. Montreal: McGill-Queen's University Press.
- McDonald, L. 1994. *The women founders of the social sciences*. Ottawa: Carleton University Press.
- McGue, M., A. Sharma, and P. Benson. 1996. The effect of common rearing on adolescent adjustment: Evidence from a U.S. adoption cohort. *Developmental Psychology* 32:604–13.
- McKinnon, A. 1993. The multi-dimensional concordance: A new tool for literary research. *Computers and the Humanities* 27:165–83.
- McLain, S., and B. Sternquist. 1991. Ethnocentric consumers: Do they “Buy American”? *Journal of International Consumer Marketing* 4:39–57.
- McLeod, R., T. Stockwell, M. Stevens, and M. Phillips. 1999. The relationship between alcohol consumption patterns and injury. *Addiction* 94(11): 1719–34.
- Mead, M. 1939. Native languages as field-work tools. *American Anthropologist* 41:189–205.
- Mead, M. 1986. Fieldwork in Pacific islands, 1925–1967. In *Women in the field: Anthropological experiences*, 2d ed., edited by P. Golde, 293–332. Berkeley: University of California Press.
- Means, B., A. Nigam, and M. Zarrow, E. F. Loftus, and M. S. Donaldson. 1989. *Autobiographical memory for health related events*. National Center for Health Statistics, Vital and Health Statistics, Series 6, No. 2. Washington, D.C: U.S. Government Printing Office.
- Meh, C. C. 1996. SOCRATES streamlines lesson observations. *Educational Leadership* 53:76–78.
- Mehta, R., and R. W. Belk. 1991. Artifacts, identity, and transition: Favorite possessions of Indians and Indian immigrants to the United States. *Journal of Consumer Research* 17:398–411.
- Melnik, T. A., S. J. Rhoades, K. R. Wales, C. Cowell, and W. S. Wolfe. 1998. Food consumption patterns of elementary schoolchildren in New York City. *Journal of the American Dietetic Association* 98:159–64.
- Mensch, I. N., and J. Henry. 1953. Direct observation and psychological tests in anthropological field work. *American Anthropologist* 55:461–80.
- Mertens, D. M. 2005. *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods*. 2d ed. Thousand Oaks, Calif.: Sage.
- Merton, R. K. 1987. The focused interview and focus groups. *Public Opinion Quarterly* 51:550–66.
- Merton, R. K., M. Fiske, and P. L. Kendall. 1956. *The focused interview: A manual of problems and procedures*. Glencoe, Ill.: The Free Press.
- Messerschmidt, D. A., ed. 1981. *Anthropologists at home in North America: Methods and issues in the study of one's own society*. New York: Cambridge University Press.
- Metzger, D. G., and G. E. Williams. 1966. Procedures and results in the study of native categories: Tselta firewood. *American Anthropologist* 68:389–407.
- Meyerhoff, B. 1989. So what do you want from us here? In *In the field: Readings on the field research experience*, edited by C. D. Smith and W. Kornblum, 83–90. New York: Praeger.
- Middlemist, R. D., E. S. Knowles, and C. F. Matter. 1976. Personal space invasion in the lavatory: Suggestive evidence for arousal. *Journal of Personality and Social Psychology* 33:541–46.

- Middlemist, R. D., E. S. Knowles, and C. F. Matter. 1977. What to do and what to report: A reply to Koocher. *Journal of Personality and Social Psychology* 35:122–24.
- Miles, M. B. 1983. *Qualitative data as an attractive nuisance: The problem of analysis*. Newbury Park, Calif.: Sage.
- Miles, M. B., and A. M. Huberman. 1994. *Qualitative data analysis*. 2d ed. Thousand Oaks, Calif.: Sage.
- Milgram, S. 1967. The small-world problem. *Psychology Today* 1:60–67.
- Milgram, S. 1969. The lost-letter technique. *Psychology Today* 3:30–33, 66–68.
- Milgram, S. 1974. *Obedience to authority*. New York: Harper & Row.
- Milgram, S. 1977a. Interpreting obedience: Error and evidence. In *The social psychology of psychological research*, edited by S. Milgram, 124–38. Reading, Mass.: Addison-Wesley.
- Milgram, S. 1977b. Ethical issues in the study of obedience. In *The individual in a social world*, edited by S. Milgram, 139–46. Reading, Mass.: Addison-Wesley.
- Milgram, S., L. Mann, and S. Harter. 1965. The lost-letter technique: A tool for social research. *Public Opinion Quarterly* 29:437–38.
- Mill, J. S. 1866. *Auguste Comte and positivism*. Philadelphia: Lippincott.
- Mill, J. S. 1869. *The subjection of women*. New York: D. Appleton & Co.
- Miller, A. S., and T. Nakamura. 1996. On the stability of church attendance patterns during a time of demographic change: 1965–1988. *The Journal for the Scientific Study of Religion* 35:275–84.
- Miller, D. C., and N. J. Salkind. 2002. *Handbook of research design and social measurement*. 6th ed. Thousand Oaks, Calif.: Sage.
- Miller, F. G., and A. Shorr 2002. Ethical assessment of industry-sponsored clinical trials: A case analysis. *CHEST* 121:1337–42.
- Miller, K. W., L. B. Wilder, F. A. Stillman, and D. M. Becker. 1997. The feasibility of a street-intercept survey method in an African-American community. *American Journal of Public Health* 87:655–58.
- Miller, L. L. 1997. Not just weapons of the weak: Gender harassment as a form of protest for Army men. *Social Psychology Quarterly* 60:32–51.
- Miller, M. L., J. Kaneko, P. Bartram, J. Marks, and D. D. Brewer. 2004. Cultural consensus analysis and environmental anthropology: Yellowfin tuna fishery management in Hawaii. *Cross-Cultural Research* 38:289–314.
- Minadeo, R. 1969. *The lyre of science: Form and meaning in Lucretius' De Rerum Natura*. Detroit: Wayne State University Press.
- Minton, A. P., and R. L. Rose. 1997. The effects of environmental concern on environmentally friendly consumer behavior: An exploratory study. *Journal of Business Research* 40:37–48.
- Mitchell, R. 1965. Survey materials collected in the developing countries: Sampling, measurement, and interviewing obstacles to intra- and international comparisons. *International Social Science Journal* 17:665–85.
- Mizes, J. S., E. L. Fleece, and C. Ross. 1984. Incentives for increasing return rates: Magnitude levels, response bias, and format. *Public Opinion Quarterly* 48:794–800.
- Momaday, N. S. 1969. *The way to Rainy Mountain*. Albuquerque: University of New Mexico Press.
- Monárrez-Espino, J., T. Greiner, and H. Martínez. 2004. Rapid qualitative assessment to design a food basket for young Tarahumara children in Mexico. *Scandinavian Journal of Nutrition* 48:4–12.
- Moran, D. 2000. *Introduction to phenomenology*. London: Routledge.
- Morgan, D. L. 1989. Adjusting to widowhood: Do social networks make it easier? *The Gerontologist* 29:101–7.

- Morgan, D. L. 1997. *Focus groups as qualitative research*. 2d ed. Thousand Oaks, Calif.: Sage.
- Morgan, D. L., and R. Krueger. 1998. *The focus group kit*, 6 vols. Thousand Oaks, Calif.: Sage.
- Morgan, L. H. 1877. *Ancient society; or, researches in the lines of human progress from savagery, through barbarism to civilization*. New York: H. Holt.
- Morgan, L. H. 1997 [1870]. *Systems of consanguinity and affinity of the human family*. Lincoln: University of Nebraska Press.
- Morin, R. 2004. Don't ask me; As fewer cooperate on polls, criticism and questions mount. *The Washington Post*, 28 October.
- Morrison, F. J., E. M. Griffith, and J. A. Frazier. 1996. Schooling and the 5 to 7 shift: A natural experiment. In *The five to seven year shift: The age of reason and responsibility*, edited by A. J. Sameroff and M. M. Haith, 161–86. Chicago: University of Chicago Press.
- Morrison, M., and T. Morrison. 1995. A meta-analytic assessment of the predictive validity of the quantitative and verbal components of the Graduate Record Examination with graduate grade point average representing the criterion of graduate success. *Educational and Psychological Measurement* 55:309–16.
- Morse, J. M. 1994. Designing funded qualitative research. In *Handbook of qualitative research*, edited by N. K. Denzin and Y. S. Lincoln, 220–35. Thousand Oaks, Calif.: Sage.
- Mulcahy, F. D., and S. Herbert. 1990. Women's formal evening wear, 1937–1982. *Journal of Social Behavior and Personality* 5(4): 481–96.
- Mumford, D. B. 1998. The measurement of culture shock. *Social Psychiatry* 33(4): 149–54.
- Murdock, G. P. 1949. *Social structure*. New York: Macmillan Co.
- Murdock, G. P. 1971. *Outline of cultural materials*. 4th rev. ed., with modifications. New Haven, Conn.: Human Relations Area Files, Inc.
- Murdock, G. P., C. S. Ford, A. E. Hudson, R. Kennedy, L. W. Simmons, and J. W. Whiting. 2004 [1961]. *Outline of cultural materials*. 5th rev. ed., modified in 2004. New Haven, Conn.: Human Relations Area Files, Inc.
- Murdock, G. P., and D. R. White. 1969. Standard cross-cultural sample. *Ethnology* 8:329–69.
- Murphy, G. P., L. B. Murphy, and T. M. Newcomb. 1937. *Experimental social psychology*. New York: Harper & Brothers.
- Murray, S. O. 1992. Components of the gay community in San Francisco. In *Gay culture in America. Essays from the field*, edited by G. Herdt, 107–46. Boston: Beacon Press.
- Murray, T. 1980. Learning to deceive. *The Hastings Center Report* 10(2): 11–14.
- Murtagh, M. 1985. The practice of arithmetic by American grocery shoppers. *Anthropology and Education Quarterly* 16:186–92.
- Mwango, E. 1986. The sources of variation in farmer adoption of government recommended technologies in the Lilongwe rural development program area of Central Malawi. Master's thesis, University of Florida.
- Nabokov, P. 1967. *Two Leggings: The making of a Crow warrior*. New York: Crowell.
- Nachman, S. R. 1984. Lies my informants told me. *Journal of Anthropological Research* 40:536–55.
- Nachmias, D., and C. Nachmias. 1976. *Research methods in the social sciences*. New York: St. Martin's Press.
- Narayan, S., and J. A. Krosnick. 1996. Education moderates some response effects in attitude measurement. *Public Opinion Quarterly* 60:58–88.
- Naroll, R. 1962. *Data quality control*. New York: The Free Press.
- National Cancer Institute. 1997. *Changes in cigarette-related disease risks and their implication for prevention and control* (microform). Developed under the general editorship of the Smoking and Tobacco Control Program (STCP), National Cancer Institute (NCI), coordinated by D. R. Shopland, edited by D. M. Burns, L. Garfinkel, and J. M. Samet, Introduc-

- tion. Bethesda, Md.: National Institutes of Health, National Cancer Institute. Summary at <http://www.nih.gov/news/pr/apr97/nci-23.htm>.
- Nave, A. 1997. Conducting a survey in a newly developed country. *Cultural Anthropology Methods Journal* 9:8–12.
- Nederhof, A. J. 1985. A survey on suicide: Using a mail survey to study a highly threatening topic. *Quality and Quantity* 19:293–302.
- Neihardt, J. 1972 [1932]. *Black Elk speaks*. New York: Pocket Books.
- Netemeyer, R. G., W. O. Bearden, and S. Sughash. 2003. Scaling procedures: Issues and applications. Thousand Oaks, Calif.: Sage.
- Newman, J. C., D. C. Des Jarlais, C. F. Turner, J. Gribble, P. Cooley, and D. Paone. 2002. The differential effects of face-to-face and computer interview modes. *American Journal of Public Health* 92:294–97.
- Newman, K. S. 1986. Symbolic dialects and generations of women: Variations in the meaning of post-divorce downward mobility. *American Ethnologist* 13:230–52.
- Nicholls, W. L., II, R. P. Baker, and J. Martin. 1997. The effect of new data collection technologies on survey data quality. In *Survey Measurement and Process Quality*, edited by L. Lyberg, P. Biemer, M. Collins, C. Dippo, E. de Leeuw, N. Schwarz, and D. Trewin, 221–48. New York: John Wiley.
- Nida, E. A. 1975. *Language structure and translation: Essays*. Stanford, Calif.: Stanford University Press.
- Nida, E. A., and C. Taber. 1969. *The theory and practice of translation*. Leiden, The Netherlands: E. J. Brill.
- Niebel, B. W. 1982. *Motion and time study*. 7th ed. Homewood, Ill.: Irwin.
- Nightingale, F. 1871. *Introductory notes on lying-in institutions*. London: Longmans, Green.
- Nisbet, R. A. 1980. *The history of the idea of progress*. New York: Basic Books.
- Nishida, H. 1999. Cognitive approach to intercultural communication based on schema theory. *International Journal of Intercultural Relations* 5:753–77.
- Nkwi, P. N. 1996. *Final report of the Cameroon Country Program*. Washington, D.C.: African Population Advisory Council.
- Nolan, J. M., and G. W. Ryan 2000. Fear and loathing at the cineplex: Gender differences in descriptions and perceptions of slasher films. *Sex Roles* 42(1–2): 39–56.
- Nooney, J. G., E. Woodrum, and T. J. Hoban. 2003. Environmental worldview and behavior: Consequences of dimensionality in a survey of North Carolinians. *Environment and Behavior* 35:763–83.
- Nordlund, S., I. Holme, and S. Tamsfoss. 1994. Randomized response estimates for the purchase of smuggled liquor in Norway. *Addiction* 89(4): 401–5.
- North, A. C., and D. J. Hargreaves. 1999. Can music move people? The effects of musical complexity and silence on waiting time. *Environment and Behavior* 31:136–49.
- Norvilitis, J. M., and H. M. Reid. 2002. Evidence for an association between gender-role identity and a measure of executive function. *Psychological Reports* 90:35–45.
- Nosek, B. A., M. Banaji, and A. G. Greenwald. 2002. Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics* 6:101–15.
- Nunnally, J. C., and I. H. Bernstein 1994. *Psychometric theory*. 3rd ed. New York: McGraw-Hill.
- Nutini, H. G., and B. Bell. 1980. *Ritual kinship*. Princeton, N.J.: Princeton University Press.
- Nyamongo, I. K. 1998. Lay people's responses to illness: An ethnographic study of anti-malaria behavior among the Abagusii of southwestern Kenya. Ph.D. diss., University of Florida.
- Oakes, J. M., and P. H. Rossi. 2003. The measurement of SES in health research: Current practice and steps toward a new approach. *Social Science and Medicine* 56:769–84.

- Oakhill, J., and A. Garnham 1996. *Mental models in cognitive science: Essays in honour of Phil Johnson-Laird*. East Sussex, England: Psychology Press.
- Oberg, K. 1960. Culture shock and the problem of adjustment in new cultural environments. *Practical Anthropology* 7:170–79.
- Oboler, R. S. 1985. *Women, power, and economic change: The Nandi of Kenya*. Stanford, Calif.: Stanford University Press.
- O'Brian, R. 1998. Stationary spot behavior and extended observation: Adapting time allocation to marketplaces. *Cultural Anthropology Methods* 10(3): 57–60.
- O'Brien, T. O., and V. Dugdale. 1978. Questionnaire administration by computer. *Journal of the Market Research Society* 20:228–37.
- Ogburn, W. F. 1930. The folk-ways of a scientific sociology. *Publication of the American Sociological Society* 25:1–10.
- Ogilvie, D. M., P. J. Stone, and E. S. Schneidman. 1966. Some characteristics of genuine versus simulated suicide notes. In *The General Inquirer: A computer approach to content analysis*, edited by P. J. Stone, D. C. Dunphy, M. S. Smith, and D. M. Ogilvie, 527–35. Cambridge: MIT Press.
- Olson, R. G. 1993. *The emergence of the social sciences, 1642–1792*. New York: Twayne Publishers.
- Opotow, S. 1994. Predicting protection: Scope of justice and the natural world. *Journal of Social Issues* 50:49–63.
- Osgood, C. E., D. J. Suci, and P. H. Tannenbaum. 1957. *The measurement of meaning*. Urbana: University of Illinois Press.
- Oskenberg, L., L. Coleman, and C. F. Cannell. 1986. Interviewers' voices and refusal rates in telephone surveys. *Public Opinion Quarterly* 50:97–111.
- Ostrander, S. A. 1980. Upper-class women: Class consciousness as conduct and meaning. In *Power structure research*, edited by G. W. Domhoff, 73–96. Beverly Hills, Calif.: Sage.
- Oths, K. S. 1994. Health care decisions of households in economic crisis: An example from the Peruvian highlands. *Human Organization* 53(3): 245–54.
- Otterbein, K. 1969. Basic steps in conducting a cross-cultural study. *Behavior Science Notes* 4:221–36.
- Owusu, M. 1978. Ethnography of Africa: The usefulness of the useless. *American Anthropologist* 80:310–34.
- Oyuela-Caycedo, A., and J. J. Vieco Albarracín. 1999. Aproximación cuantitativa a la organización social de los ticuna del trapezio amazónico colombiano. *Revista Colombiana de Antropología* 35:146–79.
- Paddock, J. 1985. Covert content in Codices Borgia and Nuttall. *Ethos* 13:358–80.
- Pagliari, C., and J. Grimshaw. 2002. Impact of group structure and process on multidisciplinary evidence-based guideline development: An observational study. *Journal of Evaluation in Clinical Practice* 8:145–53.
- Papadopoulos, F. C., E. Petridou, C. E. Frangakis, T. Farmakakis, H. Moller, and G. Rider. 2004. Switching to the Euro: Still hard to swallow. *Archives of Disease in Childhood* 89:382–83.
- Paredes, J. A. 1974. The emergence of contemporary Eastern Creek Indian identity. In *Social and cultural identity: Problems of persistence and change*. *Southern Anthropological Society Proceedings*, No. 8, edited by T. K. Fitzgerald, 63–80. Athens: University of Georgia Press.
- Paredes, J. A. 1992. "Practical history" and the Poarch Creeks: A meeting ground for anthropologist and tribal leaders. In *Anthropological research: Process and application*, edited by J. J. Poggie et al., 211–26. Albany: State University of New York Press.
- Parekh, R., and E. V. Beresin. 2001. Looking for love? Take a cross-cultural walk through the personals. *Academic Psychiatry* 25:223–33.

- Park, R. E., E. W. Burgess, and R. D. McKenzie. 1925. *The city*. Chicago: University of Chicago Press.
- Patel, J. D., P. B. Bach, and M. G. Kris. 2004. Lung cancer in U.S. women. A contemporary epidemic. *JAMA, The Journal of the American Medical Association* 291(14): 1763–68.
- Pauli, J. 2000. *Das Geplante Kind. Demographischer, Wirtschaftlicher und Sozialer Wandel in einer Mexikanischen Gemeinde*. Ph.D. thesis, University of Cologne.
- Pausewang, S. 1973. *Methods and concepts of social research in a rural and developing society*. Munich: Weltforum Verlag.
- Paxon, H. 2004. *Making modern mothers: Ethics and family planning in urban Greece*. Berkeley: University of California Press.
- Payne, S. L. 1951. *The art of asking questions*. Princeton, N.J.: Princeton University Press.
- Pearce, L. D. 2002. Integrating survey and ethnographic methods for systematic anomalous case analysis. *Sociological Methodology* 32:103–32.
- Pearson, J. 1990. Estimation of energy expenditure in Western Samoa, American Samoa, and Honolulu by recall interviews and direct observation. *American Journal of Human Biology* 2:313–26.
- Pedersen, P. B. 1993. Primal alternatives to talk therapy: The Batak “Tondi” of North Sumatra. *Counseling and Values* 37(2): 52–60.
- Pederson, J. 1987. Plantation women and children: Wage labor, adoption, and fertility in the Seychelles. *Ethnology* 26:51–62.
- Pedhazur, E. J. 1997. *Multiple regression in behavioral research: Explanation and prediction*. 3rd ed. Fort Worth, Tex.: Harcourt Brace College Publishers.
- Pellegrini, A. D. 1996. *Observing children in their natural worlds: A methodological primer*. Mahwah, N.J.: Lawrence Erlbaum.
- Pelto, G. H., P. J. Pelto, and E. Messer, eds. 1989. *Research methods in nutritional anthropology*. Tokyo: The United Nations University.
- Pelto, P. J., and G. H. Pelto. 1978. *Anthropological research: The structure of inquiry*. New York: Cambridge University Press.
- Penrod, J., D. B. Preston, R. E. Cain, and M. T. Starks. 2003. A discussion of chain referral as a method of sampling hard-to-reach populations. *Journal of Transcultural Nursing* 14:100–107.
- Perchonock, N., and O. Werner. 1969. Navajo systems of classification: Some implications of food. *Ethnology* 8:229–42.
- Pericliev, V., and R. E. Valdes-Perez. 1998. Automatic componential analysis of kinship semantics with a proposed structural solution to the problem of multiple models. *Anthropological Linguistics* 40:272–317.
- Peters, D. L., and E. J. McCormick. 1966. Comparative reliability of numerically anchored versus job-task anchored rating scales. *Journal of Applied Psychology* 50:92–96.
- Petersen, L. R., and G. V. Donnenwerth. 1997. Secularization and the influence of religion on beliefs about premarital sex. *Social Forces* 75:1071–89.
- Peterson, L., V. Johannsson, and S. G. Carlsson. 1996. Computerized testing in a hospital setting: Psychometric and psychological effects. *Computers in Human Behavior* 12:339–50.
- Peterson, M., and B. M. Johnstone. 1995. The Atwood Hall health promotion program, Federal Medical Center, Lexington, Ky.: Effects on drug-involved federal offenders. *Journal of Substance Abuse Treatment* 12:43–48.
- Peterson, R. A. 1984. Asking the age question. *Public Opinion Quarterly* 48:379–83.
- Pfeiffer, J., S. Gloyd, and L. Ramirez Li. 2001. Intrahousehold resource allocation and child growth in Mozambique: An ethnographic case-control study. *Social Science and Medicine* 53:83–97.

- Piliavin, I. M., J. Rodin, and J. A. Piliavin. 1969. Good samaritanism: An underground phenomenon? *Journal of Personality and Social Psychology* 13:289–99.
- Plattner, S. 1982. Economic decision making in a public marketplace. *American Ethnologist* 9:399–420.
- Poggie, J. J. 1972. Toward quality control in key informant data. *Human Organization* 31:23–30.
- Poggie, J. J. 1979. Small-scale fishermen's beliefs about success and development: A Puerto Rican case. *Human Organization* 38:6–11.
- Pollner, M., and R. E. Adams. 1997. The effect of spouse presence on appraisals of emotional support and household strain. *Public Opinion Quarterly* 61:615–26.
- Pollock, P. H., III. 1994. Issues, values and critical moments: Did "Magic" Johnson transform public opinion on AIDS? *American Journal of Political Science* 38:426–46.
- Posey, D. A. 2004. *Indigenous knowledge and ethics: A Darrell Posey reader*. New York: Routledge.
- Powdermaker, H. 1966. *Stranger and friend: The way of an anthropologist*. New York: Norton.
- Prebisch, R. 1984. *Power relations and market laws*. Notre Dame, Ind.: The Helen Kellogg Institute for International Studies, University of Notre Dame.
- Prebisch, R. 1994. Latin American periphery in the global system of capitalism. In *Paradigms in economic development: Classic perspectives, critiques, and reflections*, edited by R. Kanth, 165–76. Armonk, N.Y.: M. E. Sharpe.
- Presser, S., and L. Stinson. 1998. Data collection mode and social desirability bias in self-reported religious attendance. *American Sociological Review* 63:137–45.
- Presser, S., and S. Zhao. 1992. Attributes of questions and interviewers as correlates of interviewing performance. *Public Opinion Quarterly* 56:236–40.
- Price, D. H. 1993. The evolution of irrigation in Egypt's Fayoum Oasis: State, village and conveyance loss. Ph.D. diss., University of Florida.
- Price, D. H. 1995. Water theft in Egypt's Fayoum oasis: Emics, etics, and the illegal. In *Science, materialism, and the study of culture*, edited by M. F. Murphy and M. L. Margolis, 96–110. Gainesville: University Press of Florida.
- Price, J. H., T. L. Kandakai, S. Casler, S. Everett, and D. Smith. 1994. African-American adults' perceptions of guns and violence. *Journal of the National Medical Association* 86:426–32.
- Price, L. 1987. Ecuadorian illness stories. In *Cultural models in language and thought*, edited by D. Holland and N. Quinn, 313–42. Cambridge: Cambridge University Press.
- Proctor, S. P., R. F. White, T. G. Robins, and D. Echeverria. 1996. Effect of overtime work on cognitive function in automotive workers. *Scandinavian Journal of Work, Environment and Health* 22:124–32.
- Psathas, G., ed. 1979. *Everyday language*. New York: Irvington Publishers.
- Psathas, G., and T. Anderson. 1990. The 'practices' of transcription in conversation analysis. *Semiotica* 78:75–99.
- Puccio, E. M., J. B. McPhillips, E. Barrett-Connor, and T. G. Ganiats. 1990. Clustering of atherogenic behaviors in coffee drinkers. *American Journal of Public Health* 80:1310–13.
- Qin, L., and Y. Su. 2003. Revision of Bem Sex Role Inventory. *Chinese Mental Health Journal* 17:550–53.
- Quandt, S. A., and C. Ritenbaugh. 1986. *Training manual in nutritional anthropology*. Washington, D.C.: American Anthropological Association.
- Quandt, S. A., M. Z. Vitolins, K. M. DeWalt, and G. Roos. 1997. Meal patterns of older adults in rural communities: Life course analysis and implications for undernutrition. *Journal of Applied Gerontology* 16:152–71.
- Quételet, A. 1969 [1835]. *Physique sociale, ou, essai sur le développement des facultés de l'homme*. Paris: J.-B. Bailliere et fils. Reprinted in trans. in 1969 from the 1842 ed. as A

- treatise on man and the development of his faculties*. Gainesville, Fla.: Scholars' Facsimiles and Reprints.
- Quinn, N. 1978. Do Mfantse fish sellers estimate probabilities in their heads? *American Ethnologist* 5:206–26.
- Quinn, N. 1982. "Commitment" in American marriage: A cultural analysis. *American Ethnologist* 9:755–98.
- Quinn, N. 1987. Convergent evidence for a cultural model of American marriage. In *Cultural models in language and thought*, edited by N. Quinn and D. Holland, 173–92. Cambridge: Cambridge University Press.
- Quinn, N. 1992. The motivational force of self-understanding: Evidence from wives' inner conflicts. In *Human motives and cultural models*, edited by R. G. D'Andrade and C. Stauss, 90–126. New York: Cambridge University Press.
- Quinn, N. 1996. Culture and contradiction: The case of Americans reasoning about marriage. *Ethos* 24:391–425.
- Quinn, N. 1997. Research on the psychodynamics of shared task solutions. In *A cognitive theory of cultural meaning*, edited by C. Strauss and N. Quinn, 189–209. Cambridge: Cambridge University Press.
- Radin, P. 1966 [1933]. *The method and theory of ethnology*. New York: Basic Books.
- Radin, P. 1983 [1920]. *The autobiography of a Winnebago Indian*. University of California Publications in American Archaeology and Ethnology, Vol. 16, No. 7. Berkeley: University of California Press.
- Ragin, C. C. 1987. *The comparative method: Moving beyond qualitative and quantitative strategies*. Berkeley: University of California Press.
- Ragin, C. C. 1994. Introduction to qualitative comparative analysis. In *The comparative political economy of the welfare state*, edited by T. Janowski and A. M. Hicks, 299–317. Cambridge: Cambridge University Press.
- Ragin, C. C. 1998. The logic of qualitative comparative analysis. *International Review of Social History* 43(supplement 6): 105–24.
- Rand Corporation. 1965. *A million random digits with 100,000 normal deviates*. Glencoe, Ill.: The Free Press.
- Rantala, K., and E. Hellstrom. 2001. Qualitative comparative analysis and a hermeneutic approach to interview data. *International Journal of Social Research Methodology* 4(2): 87–100.
- Rappaport, R. 1990. Forward. In *Surviving fieldwork*, edited by N. Howell, vii–viii. Washington, D.C.: American Anthropological Association.
- Rathje, W. L. 1979. Trace measures. Garbage and other traces. In *Unobtrusive measurement today*, edited by L. Sechrest, 75–91. San Francisco: Jossey-Bass.
- Rathje, W. L. 1984. The garbage decade. In *Household refuse analysis: Theory, method, and applications in social science*, edited by W. L. Rathje and C. K. Rittenbaugh, 9–29. *American Behavioral Scientist* 28(1, entire issue).
- Rathje, W. L., and C. Murphy. 1992. Garbage demographics. *American Demographics* 14(5): 50–54.
- Reboussin, D. 1995. From Affiniam-Boutem to Dakar: Migration from the Casamance, life in the urban environment of Dakar and the resulting evolutionary changes in local Diola organizations. Ph.D. diss., University of Florida.
- Reed, T. W., and R. J. Stimson, eds. 1985. *Survey interviewing. Theory and techniques*. Sydney: Allen & Unwin.
- Reese, S. D., W. A. Danielson, P. J. Shoemaker, Tsan-Kuo Chang, and Hwei-ling Hsu. 1986. Eth-

- nicity-of-interviewer effects among Mexican-American and Anglos. *Public Opinion Quarterly* 50:563–72.
- Reiss, A. J., Jr. 1971. *The police and the public*. New Haven, Conn.: Yale University Press.
- Reiss, N. 1985. *Speech act taxonomy as a tool for ethnographic description: An analysis based on videotapes of continuous behavior in two New York households*. Philadelphia: John Benjamins.
- Remland, M., T. S. Jones, and H. Brinkman. 1995. Interpersonal distance, body orientation, and touch: Effects of culture, gender, and age. *The Journal of Social Psychology* 135:281–97.
- Repp, A. C., G. S. Nieminen, E. Olinger, and R. Brusca. 1988. Direct observation: Factors affecting the accuracy of observers. *Exceptional Children* 55:29–36.
- Ricci, J. A., N. W. Jerome, N. Megally, and O. Galal. 1995. Assessing the validity of information recall: Results of a time use pilot study in peri-urban Egypt. *Human Organization* 54:304–8.
- Richards, A. I. 1939. *Land, labour and diet in northern Rhodesia: An economic study of the Bemba tribe*. London: Oxford University Press.
- Richardson, J., and A. L. Kroeber. 1940. Three centuries of women's dress fashions: A quantitative analysis. *Anthropological Records* 5(2): 111–53.
- Richardson, L. 1988. Secrecy and status: The social construction of forbidden relationships. *American Sociological Review* 53:209–19.
- Rickford, J. R., and R. J. Rickford. 2000. *Spoken soul: The story of black English*. New York: Wiley.
- Ricoeur, P. 1979. *Main trends in philosophy / Paul Ricoeur*. New York: Holmes & Meier.
- Ricoeur, P. 1981. *Hermeneutics and the human sciences: Essays on language, action, and interpretation*. New York: Cambridge University Press.
- Ricoeur, P. 1991. *From text to action*. Evanston, Ill.: Northwestern University Press.
- Rindfleisch, A., J. E. Burroughs, and F. Denton. 1997. Family structure, materialism, and compulsive consumption. *Journal of Consumer Research* 23:312–25.
- Rittenbaugh, C. K., and G. G. Harrison. 1984. Reactivity of garbage analysis. In *Household refuse analysis: Theory, method, and applications in social science*, edited by W. L. Rathje and C. K. Rittenbaugh, 51–70. *American Behavioral Scientist* 28(1, entire issue).
- Rivers, W.H.R. 1906. *The Todas*. New York: Macmillan
- Rivers, W.H.R. 1910. The genealogical method of anthropological inquiry. *Sociological Review* 3:1–12.
- Rivers, W.H.R. 1968 [1914]. *Kinship and social organization*. New York: Humanities Press.
- Robbins, M. C., A. V. Williams, P. L. Killbride, and R. B. Pollnac. 1969. Factor analysis and case selection in complex societies. *Human Organization* 28:227–34.
- Roberts, J. M., and G. E. Chick. 1979. Butler County eight-ball: A behavioral space analysis. In *Sports, games, and play: Social and psychological viewpoints*, edited by J. H. Goldstein, 65–100. Hillsdale, N.J.: Lawrence Erlbaum.
- Roberts, J. M., and S. Nattrass. 1980. Women and trapshooting: Competence and expression in a game of physical skill with chance. In *Play and culture*, edited by H. B. Schwartzman, 262–90. West Point, N.Y.: Leisure Press.
- Roberts, J. M., T. V. Golder, and G. E. Chick. 1980. Judgment, oversight, and skill: A cultural analysis of P-3 pilot error. *Human Organization* 39:5–21.
- Roberts, L. J., D. A. Luke, J. Rappaport, E. Seidman, P. A. Toro, and T. M. Reischl. 1991. Charting uncharted terrain: A behavioral observation system for mutual help groups. *American Journal of Community Psychology* 19:715–37.
- Robinson, D., and S. Rhode. 1946. Two experiments with an anti-Semitism poll. *Journal of Abnormal and Social Psychology* 41:136–44.

- Robinson, M. B., and C. E. Robinson. 1997. Environmental characteristics associated with residential burglaries of student apartment complexes. *Environment and Behavior* 29:657–75.
- Robinson, W. S. 1950. Ecological correlations and the behavior of individuals. *American Sociological Review* 15:351–57.
- Rogoff, B. 1978. Spot observation: An introduction and examination. *Quarterly Newsletter of the Institute for Comparative Human Development* 2:21–26.
- Rohner, R. 1966. Franz Boas, ethnographer of the Northwest Coast. In *Pioneers of American anthropology*, edited by J. Helm, 149–212. Seattle: University of Washington Press.
- Rohner, R. 1969. *The ethnography of Franz Boas*. Chicago: University of Chicago Press.
- Rohner, R., B. R. DeWalt, and R. C. Ness. 1973. Ethnographer bias in cross-cultural research. *Behavior Science Notes* 8:275–317.
- Romney, A. K. 1989. Quantitative models, science and cumulative knowledge. *Journal of Quantitative Anthropology* 1:153–223.
- Romney, A. K., and R. G. D'Andrade, eds. 1964. Cognitive aspects of English kin terms. In *Transcultural studies in cognition*. *American Anthropologist* 66 (3, part 2, entire issue): 146–70.
- Romney, A. K., R. N. Shepard, and S. B. Nerlove, eds. 1972. *Multidimensional scaling*. Vol. 2, *Applications*. New York: Seminar Press.
- Romney, A. K., S. C. Weller, and W. H. Batchelder. 1986. Culture as consensus: A theory of culture and informant accuracy. *American Anthropologist* 88:313–38.
- Roscigno, V. J., and R. Hodson. 2004. The organizational and social foundations of worker resistance. *American Sociological Review* 69:14–39.
- Rose, M. G. 1997. *Translation and literary criticism: Translation as analysis*. Manchester, England: St. Jerome.
- Rosenberg, M. 1968. *The logic of survey analysis*. New York: Basic Books.
- Rosenberg, S. D., P. P. Schnurr, and T. E. Oxman. 1990. Content analysis: A comparison of manual and computerized systems. *Journal of Personality Assessment* 54:298–310.
- Rosenhan, D. L. 1973. On being sane in insane places. *Science* 179:250–58.
- Rosenhan, D. L. 1975. The contextual nature of psychiatric diagnosis. *Journal of Abnormal Psychology* 84:462–74.
- Rosenthal, R. 2000. Expectancy effects. In *Encyclopedia of psychology*, Vol. 3, edited by A. E. Kazdin, 294–96. Washington, D.C.: American Psychological Association.
- Rosenthal, R. 2002. Experimenter and clinician effects in scientific inquiry and clinical practice. *Prevention and Treatment* 5 [electronic journal].
- Rosenthal, R., and L. Jacobson. 1968. *Pygmalion in the classroom*. New York: Holt, Rinehart & Winston.
- Rosenthal, R., and D. B. Rubin. 1978. Interpersonal expectancy effects: The first 345 studies. *The Behavioral and Brain Sciences* 3:377–415.
- Ross, M. W., B. R. Simon Rosser, J. Stanton, and J. Konstan. 2004. Characteristics of Latino men who have sex with men on the Internet who complete and drop out of an Internet-based sexual behavior survey. *AIDS Education and Prevention* 16:526–37.
- Rossi, P. H., and R. Berk. 1997. *Just punishments: Federal guidelines and public views compared*. New York: Aldine de Gruyter.
- Rossi, P. H., and S. L. Nock. 1982. *Measuring social judgments: The factorial survey approach*. Newbury Park, Calif.: Sage.
- Roth, P. L., and C. S. BeVier. 1998. Response rates in HRM/OB survey research: Norms and correlates, 1990–1994. *Journal of Management* 24(1): 97–117.
- Rothschild, R. F. 1981. What happened in 1780? *Harvard Magazine* 83 (January–February): 20–27.

- Rotton, J., and M. Shats. 1996. Effects of state humor, expectancies, and choice on postsurgical mood and self-medication: A field experiment. *Journal of Applied Social Psychology* 26:1775–94.
- Rotton, J., M. Shats, and R. Standers. 1990. Temperature and pedestrian tempo. Walking without awareness. *Environment and Behavior* 22:650–74.
- Rousseau, J.-J. 1988 [1762]. *On the social contract*. Translated and edited by D. A. Cress. Indianapolis: Hackett Publishing.
- Royal Anthropological Institute of Great Britain and Ireland (RAI). 1951. *Notes and queries on anthropology*. 6th ed., rev. London: Routledge & K. Paul.
- Rozeboom, W. 1960. The fallacy of the null-hypothesis significance test. *Psychological Bulletin* 67:416–28.
- Rubin, D. B. 1976. Inference and missing data. *Biometrika* 63:581–92.
- Rushforth, S. 1982. A structural semantic analysis of Bear Lake Athapaskan kinship classification. *American Ethnologist* 9:559–77.
- Rutledge, E. M. 1990. Black parent-child relations: Some correlates. *Journal of Comparative Family Studies* 21:369–78.
- Ryan, G. W. 1995. Medical decision making among the Kom of Cameroon: Modeling how characteristics of illnesses, patients, caretakers, and compounds affect treatment choice in a rural community. Ph.D. diss., University of Florida.
- Ryan, G. W., and H. R. Bernard. 2000. Data management and analysis methods. In *Handbook of qualitative research*, 2d ed., edited by N. Denzin and Y. Lincoln, 769–802. Thousand Oaks, Calif.: Sage.
- Ryan, G. W., and H. R. Bernard. 2003. Techniques to identify themes. *Field Methods* 15:85–109.
- Ryan, G. W., and H. Martínez. 1996. Can we predict what mothers do? Modeling childhood diarrhea in rural Mexico. *Human Organization* 55:47–57.
- Ryan, G. W., and T. Weisner. 1996. Analyzing words in brief descriptions: Fathers and mothers describe their children. *Cultural Anthropology Methods Journal* 8:13–16.
- Rynkiewicz, M., and J. P. Spradley, eds. 1976. *Ethics and anthropology: Dilemmas in fieldwork*. New York: John Wiley.
- Sackett, R. 1996. Time, energy, and the indolent savage. A quantitative cross-cultural test of the primitive affluence hypothesis. Ph.D. diss., University of California, Los Angeles.
- Sacks, H. 1992. *Lectures on conversation*. Cambridge, Mass.: Basil Blackwell.
- Sacks, H., E. A. Schegloff, and G. Jefferson. 1974. A simplest systematics for the organization of turn-taking conversation. *Language* 50:696–735.
- Sahlins, M. 1972. *Stone Age economics*. Chicago: Aldine.
- Salant, P., and D. Dillman. 1994. *How to conduct your own survey*. New York: Wiley.
- Salganik, M. J., and D. D. Heckathorn. 2004. Sampling and estimation in hidden populations using respondent-driven sampling. *Sociological Methodology* 34:193–239.
- Salinas Pedraza, J. 1984. *Emografía del Otomí*. Mexico City: Instituto Nacional Indigenista.
- Salinas Pedraza, J., and H. R. Bernard 1978. *Rc Hnynhyu. The Otomí*. Vol. 1, *Geography and fauna*. Albuquerque: University of New Mexico Press.
- Salmon, M. H. 1997. Ethical considerations in anthropology and archaeology; or, Relativism and justice for all. *Journal of Anthropological Research* 53:47–63.
- Sanchez, M. E., and G. Morchio. 1992. Probing “Don’t Know” answers: Effects on survey estimates and variable relationships. *Public Opinion Quarterly* 56:454–74.
- Sandall, R. 1999. Herskovits’ last day in Dahomey. *Anthropology Today* 15(6): 18–20.
- Sands, K. 1980. Preface. In *Autobiography of a Yaqui poet*, edited by R. Savala, vii–ix. Tucson: University of Arizona Press.
- Sanjek, R. 1990. *Fieldnotes*. Ithaca, N.Y: Cornell University Press.

- Sankoff, G. 1971. Quantitative analysis of sharing and variability in a cognitive model. *Ethnology* 10:389–408.
- Sarton, G. 1935. Quételet (1796–1874). *Isis* 23:6–24.
- SAUS 1947. Statistical abstract of the United States. Washington, DC: U.S. Census Bureau.
- SAUS 2004–2005. Statistical abstract of the United States. <http://www.census.gov/prod/www/statistical-abstract-04.html>.
- Savala, R. 1980. *Autobiography of a Yaqui poet*. Tucson: University of Arizona Press.
- Scaglione, R. 1986. The importance of nighttime observations in time allocation studies. *American Ethnologist* 13:537–45.
- Schaeffer, N. C. 2000. Asking questions about threatening topics: A selective overview. In *The science of self-report: Implications for research and practice*, edited by A. A. Stone, J. S. Turkkan, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman, and V. S. Cain, 105–22. Mahwah, N.J.: Lawrence Erlbaum.
- Schaeffer, N. C., and S. Presser. 2003. The science of asking questions. *Annual Review of Sociology* 29:65–88.
- Schank, R. C., and R. P. Abelson. 1977. *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Hillsdale, N.J.: Lawrence Erlbaum.
- Scheers, N. J., and C. M. Dayton. 1987. Improved estimation of academic cheating behavior using the randomized response technique. *Research in Higher Education* 26:61–69.
- Schegloff, E. A. 1972. Sequencing in conversational opening. In *Directions in sociolinguistics*, edited by J. Gumpertz, 346–80. New York: Holt, Reinhart and Winston.
- Schegloff, E. A., and H. Sacks. 1973. Opening up closings. *Semiotica* 7:289–327.
- Schensul, S. L., J. Schensul, G. Oodit, U. Bowan, and S. Ragobur. 1994. Sexual intimacy and changing lifestyles in an era of AIDS. *Reproductive Health Matters* No. 3(May): 83–92.
- Scheper-Hughes, N. 1983. Introduction: The problem of bias in androcentric and feminist anthropology. In *Confronting problems of bias in feminist anthropology*, edited by N. Scheper-Hughes, 109–16. *Women's Studies* 10 (special issue).
- Scheper-Hughes, N. 1992. *Death without weeping. The violence of everyday life in Brazil*. Berkeley: University of California Press.
- Scherer, S. E. 1974. Proxemic behavior of primary-school children as a function of the socioeconomic class and subculture. *Journal of Personality and Social Psychology* 29:800–805.
- Schiller, F.C.S. 1969 [1903]. *Humanism. Philosophical essays*. Freeport, N.Y.: Books for Libraries Press.
- Schlegel, A., and H. Barry, III. 1986. The cultural consequences of female contribution to subsistence. *American Anthropologist* 88:142–50.
- Schober, M. F., and F. G. Conrad. 1997. Does conversational interviewing reduce survey measurement error? *Public Opinion Quarterly* 61:576–602.
- Schoenberg, N. E. 1997. A convergence of health beliefs: An “ethnography of adherence” of African-American rural elders with hypertension. *Human Organization* 56:174–81.
- Schoenberg, N. E. 2000. Patterns, factors, and pathways contributing to nutritional risk among rural African American elders. *Human Organization* 59:235–44.
- Schuman, H., and S. Presser. 1979. The open and closed question. *American Sociological Review* 44:692–712.
- Schuman, H., and S. Presser. 1981. *Questions and answers in attitude surveys*. San Diego: Academic Press.
- Schuster, J. A. 1977. *Descartes and the scientific revolution*. Princeton, N.J.: Princeton University Press.
- Schutz, A. 1962. *Collected papers I: The problem of social reality*. The Hague: Martinus Nijhoff.

- Schwartz, N. A. 2004. Childhood asthma on the northern Mexico border. *Medical Anthropology Quarterly* 18:214–29.
- Schwarz, B., H-P. Bischof, and M. Kunze. 1994. Coffee, tea, and lifestyle. *Preventive Medicine* 23:377–84.
- Schwarz, N. 1966. *Cognition and communication: Judgment biases, research methods, and the logic of conversation*. Hillsdale, N.J.: Lawrence Erlbaum.
- Schwarz, N. 1999. Self-reports. How the questions shape the answers. *American Psychologist* 54:93–105.
- Schwarz, N., H.-J. Hippler, B. Deutsch, and F. Strack. 1985. Response scales: Effects of category range on reported behavior and comparative judgments. *Public Opinion Quarterly* 49:388–95.
- Schwarz, N., B. Knäuper, H.-J. Hippler, E. Noelle-Neumann, and L. Clark. 1991. Rating scales: Numeric values may change the meaning of scale labels. *Public Opinion Quarterly* 55:570–82.
- Schweizer, T. 1991. The power struggle in a Chinese community, 1950–1980: A social network analysis of the duality of actors and events. *Journal of Quantitative Anthropology* 3:19–44.
- Schweizer, T. 1996. Actor and event orderings across time: Lattice representation and Boolean analysis of the political disputes in Chen Village, China. *Social Networks* 18:247–66.
- Science. 1972. The Brawling Bent. 24 March:1346–47.
- Scollon, C. N., E. Diener, and S. Oishi. 2005. An experience sampling and cross-cultural investigation of the relation between pleasant and unpleasant affect. *Cognition and Emotion* 19:27–52.
- Scribner, S., and M. Cole. 1981. *The psychology of literacy*. Cambridge: Harvard University Press.
- Scrimshaw, N., and G. R. Gleason, eds. 1992. *RAP. Rapid Assessment Procedures. Qualitative methodologies for planning and evaluation of health related programmes*. Boston: International Nutrition Foundation for Developing Countries.
- Scrimshaw, S.C.M., and E. Hurtado. 1987. *Rapid assessment procedures for nutrition and primary health care: Anthropological approaches to improving programme effectiveness*. Los Angeles: UCLA Latin American Center Publications, University of California.
- Scudder, T., and E. Colson. 2002. Long-term research in Gwembe Valley, Zambia. In *Chronicling cultures: Long-term field research in anthropology*, edited by R. V. Kemper and A. P. Royce, 197–238. Walnut Creek, Calif.: AltaMira.
- Sechrest, L., and L. Flores. 1969. Homosexuality in the Philippines and the United States: The handwriting on the wall. *Journal of Social Psychology* 79:3–12.
- Sechrest, L., and M. Phillips. 1979. Unobtrusive measures: An overview. In *Unobtrusive measurement today*, edited by L. Sechrest, 1–17. San Francisco: Jossey-Bass.
- Sexton, J. D., ed. 1981. Background of the story. In *I. Bizarro Ujpán, Son of Tecún Umán: A Maya Indian tells his life story*, 3–20. Tucson: University of Arizona Press.
- Sexton, J. D., and I. Bizarro Ujpán. 2001. *Joseño: Another Mayan voice speaks from Guatemala*. Albuquerque: University of New Mexico Press.
- Sharff, J. W. 1979. Patterns of authority in two urban Puerto Rican households. Ph.D. diss., Columbia University.
- Sharpe, R. 1998. EEOC backs away from filing race-bias suit. *The Wall Street Journal*, 24 June: A4
- Shaw, C. R. 1930. *The jack-roller. A delinquent boy's own story*. Chicago: University of Chicago Press.
- Shaw, S. 1992. Dereifying family leisure: An examination of women's and men's everyday experiences and perceptions of family time. *Leisure Studies* 14:271–86.

- Sheatsley, P. B. 1983. Questionnaire construction and item wording. In *Handbook of survey research*, edited by P. H. Rossi, J. D. Wright, and A. B. Anderson, 195–230. New York: Wiley.
- Sheets, J. W. 1982. Nonleptokurtic marriage distances on Colonsay and Jura. *Current Anthropology* 23:105–6.
- Shelley, G. A. 1992. The social networks of people with end-stage renal disease: Comparing hemodialysis and peritoneal dialysis patients. Ph.D. diss., University of Florida.
- Shelley, G. A., H. R. Bernard, and P. D. Killworth. 1990. Information flow in social networks. *Journal of Quantitative Anthropology* 2:201–25.
- Shelley, G. A., H. R. Bernard, P. D. Killworth, E. C. Johnsen, and C. McCarty 1995. Who knows your HIV status? What HIV + patients and their network members know about each other. *Social Networks* 17:189–217.
- Sherman, E., A. Mathur, and R. B. Smith. 1997. Store environment and consumer purchase behavior: Mediating role of consumer emotions. *Psychology and Marketing* 14:361–78.
- Sherry, J. F., Jr. 1995. *Contemporary marketing and consumer behavior: An anthropological sourcebook*. Thousand Oaks, Calif.: Sage.
- Sherzer, J. 1990. *Verbal art in San Blas*. Cambridge: Cambridge University Press.
- Sherzer, J. 1994. The Kuna and Columbus: Encounters and confrontations of discourse. *American Anthropologist* 96:902–25.
- Sherzer, J., and G. Urban. 1986. Introduction. In *Native South American discourse*, edited by J. Sherzer and G. Urban, 1–14. Berlin: Mouton.
- Sherzer, J., and A. Woodbury. 1987. Introduction. In *Native American discourse: Poetics and rhetoric*, edited by J. Sherzer and A. Woodbury, 1–16. Cambridge: Cambridge University Press.
- Shostak, M. 1981. *Nisa: The life and words of a !Kung woman*. Cambridge, Mass.: Harvard University Press.
- Shweder, R., and R. G. D'Andrade. 1980. The systematic distortion hypothesis. In *Fallible judgment in behavioral research*, edited by R. Shweder, 37–58. San Francisco: Jossey-Bass.
- Simmons, L. W., ed. 1942. *Sun Chief: The autobiography of a Hopi Indian*. New Haven, Conn.: Yale University Press.
- Simpson, J. A., and S. W. Gangestad. 1991. Individual differences in sociosexuality: Evidence for convergent and discriminant validity. *Journal of Personality and Social Psychology* 60:870–83.
- Sin, L. Y. M., and O. H. M. Yau. 2004. Female role orientation of Chinese women: Conceptualization and scale development. *Psychology and Marketing* 21:1033–58.
- Sirken, M. G. 1972. *Designing forms for demographic surveys*. Chapel Hill: Laboratories for Population Statistics, Manual Series, No. 3. The Department of Biostatistics, School of Public Health, The Carolina Population Center, University of North Carolina.
- Skinner, B. F. 1957. *Verbal behavior*. New York: Appleton-Century-Crofts.
- Skruppy, M. 1993. Activities of daily living evaluations: Is there a difference in what the patient reports and what is observed? *Physical and Occupational Therapy in Geriatrics* 11:13–25.
- Smith, C. A., and S. Stillman. 2002. What do women want? The effects of gender and sexual orientation on the desirability of physical attributes in the personal ads of women. *Sex Roles* 46(9–10): 337–42.
- Smith, C. D., and W. Kornblum, eds. 1996. *In the field. Readings on the research experience*. Westport, Conn.: Praeger.
- Smith, C. J., R. G. Nelson, S. A. Hardy, E. M. Manahan, P. H. Bennett, and W. C. Knowler. 1996. Survey of the diet of Pima Indians using quantitative food frequency assessment and 24-hour recall. *Journal of the American Dietetic Association* 96:778–84.

- Smith, L. D. 1986. *Behaviorism and logical positivism*. Stanford, Calif.: Stanford University Press.
- Smith, M. L., and G. V. Glass. 1977. Meta-analysis of psychotherapy outcome studies. *American Psychologist* 32:752–60.
- Smith, P. C., L. M. Kendall, and C. L. Hulin. 1969. *The measurement of satisfaction in work and retirement: A strategy for the study of attitudes*. Chicago: Rand McNally.
- Smith, R. 1997. *The Norton history of the human sciences*. New York: W.W. Norton.
- Smith, T. W. 1987. That which we call welfare by any other name would smell sweeter. An analysis of the impact of question wording on response patterns. *Public Opinion Quarterly* 51:75–83.
- Smith, T. W. 1997. The impact of the presence of others on a respondent's answers to questions. *International Journal of Public Opinion Research* 9:33–47.
- Smith, T. W. 1998. A review of church attendance measures. *American Sociological Review* 63:131–36.
- Smits, J., W. Ultee, and J. Lammers. 1998. Educational homogamy in 65 countries: An explanation of differences in openness using country-level explanatory variables. *American Sociological Review* 63:264–85.
- Snider, J. G., and C. E. Osgood., eds. 1969. *Semantic differential technique*. Chicago: Aldine.
- Soifer, V. 1994. *Lysenko and the tragedy of Soviet science*. Translated by L. Grulioiw and R. Grulioiw. New Brunswick, N.J.: Rutgers University Press.
- Sokolovsky, J., C. Cohen, D. Berger, and J. Geiger. 1978. Personal networks of ex-mental patients in a Manhattan SRO hotel. *Human Organization* 37:5–15.
- Sokolowski, R. 2000. *Introduction to phenomenology*. New York: Cambridge University Press.
- Solomon, C. 1993. Transportation: Self-service at gas stations includes paying. *The Wall Street Journal*, 4 August:B1.
- Sousa Campos, L. de, E. Otta, and J. de Oliveira Siqueira. 2002. Sex differences in mate selection strategies: Content analyses and responses to personal advertisements in Brazil. *Evolution and Human Behavior* 23:395–406.
- South, S. J., and K. D. Crowder. 2000. The declining significance of neighborhoods? Marital transitions in community context. *Social Forces* 78(3): 1067–99.
- Spector, P. E. 1992. *Summated rating scale construction*. Newbury Park, Calif.: Sage.
- Spence, J. T. 1991. Do the BSRI and the PAQ measure the same or different concepts? *Psychology of Women Quarterly* 15:141–65.
- Spence, J. T., and R. L. Helmreich. 1978. *Masculinity and femininity: Their psychological dimensions, correlates, and antecedents*. Austin: University of Texas Press.
- Spence, J. T., R. L. Helmreich, and J. Stapp. 1974. The Personal Attributes Questionnaire: A measure of sex-role stereotypes and masculinity-femininity. *Catalog of Selected Documents in Psychology* 4:43–44.
- Spradley, J. P. 1972. Adaptive strategies of urban nomads. In *Culture and cognition: Rules, maps, and plans*, edited by J. P. Spradley, 235–78. New York: Chandler.
- Spradley, J. P. 1979. *The ethnographic interview*. New York: Holt, Rinehart & Winston.
- Spradley, J. P. 1980. *Participant observation*. New York: Holt, Rinehart & Winston.
- Spring, M., J. Westermeyer, L. Halcon, K. Savik, C. Robertson, D. R. Johnson, J. N. Butcher, and J. Jaranson. 2003. Sampling in difficult to access refugee and immigrant communities. *Journal of Nervous and Mental Disease* 191:813–19.
- Sproull, L. S. 1981. Managing education programs: A micro-behavioral analysis. *Human Organization* 40:113–22.
- Squire, P. 1988. Why the 1936 “Literary Digest” poll failed. *Public Opinion Quarterly* 52:125–33.

- Srinivas, M. N. 1979. The fieldworker and the field: A village in Karnataka. In *The fieldworker and the field*, edited by M. N. Srinivas, A. M. Shah, and E. A. Ramaswamy, 19–28. Delhi: Oxford University Press.
- Staats, S., D. Miller, M. J. Carnot, K. Rada, and J. Turner. 1996. The Miller-Rada Commitment to Pets Scale. *Anthrozoos* 9(2–3): 88–94.
- Stearman, A. M. 1989. Yuquí foragers in the Bolivian Amazon: Subsistence strategies, prestige, and leadership in an acculturating society. *Journal of Anthropological Research* 45:219–44.
- Stein, J., S. A. Fox, and P. J. Murata. 1991. The influence of ethnicity, socioeconomic status, and psychological barriers on the use of mammography. *Journal of Health and Social Behavior* 32:101–13.
- Stein, R. H. 1987. *The synoptic problem*. Grand Rapids, Mich.: Baker Book House.
- Stemmer, N. 1990. Skinner's "Verbal Behavior," Chomsky's review, and mentalism. *Journal of the Experimental Analysis of Behavior* 54:307–16.
- Stephenson, J. B., and L. S. Greer. 1981. Ethnographers in their own cultures: Two Appalachian cases. *Human Organization* 30:333–43.
- Sterk, C. E. 1989. Prostitution, drug use, and AIDS. In *In the field: Readings on the field research experience*, edited by C. D. Smith and W. Kornblum, 91–100. New York: Praeger.
- Sterk, C. E. 1999. *Tricking and tripping: Prostitution in the era of AIDS*. Putnam Valley, N.Y.: Social Change Press.
- Stevens, S. S. 1946. On the theory of measurement. *Science* 103:677–80.
- Steward, J. H. 1949. Cultural causality and law: A trial formulation of early civilization. *American Anthropologist* 51:1–27.
- Steward, J. H. 1955. *Theory of culture change; the methodology of multilineal evolution*. Urbana: University of Illinois Press.
- Stewart, D. W., and P. N. Shamdasani. 1990. *Focus groups: Theory and practice*. Newbury Park, Calif.: Sage.
- Stewart, M. A. 1984. What is a successful doctor-patient interview? A study of interactions and outcomes. *Social Science and Medicine* 19:167–75.
- Stocking, G. W., Jr. 1983. *Observers observed: Essays on ethnographic fieldwork*. Madison: University of Wisconsin Press.
- Stone, P. J., D. C. Dunphy, M. S. Smith, and D. M. Ogilvie. 1966. *The General Inquirer: A computer approach to content analysis*. Cambridge: MIT Press.
- Storer, N. W. 1966. *The social system of science*. New York: Holt, Rinehart & Winston.
- Storosum, J. G., B. J. van Zwieten, and T. Wohlfarth. 2003. Suicide risk in placebo vs. active treatment in placebo-controlled trials for schizophrenia. *Archives of General Psychiatry* 60:365–68.
- Stouffer, S. A., E. A. Suchman, L. C. DeVinney, S. A. Star, and R. M. Williams. 1949. *The American soldier: Adjustment during army life*. Princeton, N.J.: Princeton University Press.
- Strathern, A. 1979. *Ongka: A self-account by a New Guinea big man*. New York: St. Martin's.
- Strauss, A., and J. Corbin. 1990. *Basics of qualitative research. Grounded theory procedures and techniques*. Thousand Oaks, Calif.: Sage.
- Strauss, C. 1992. What makes Tony run? Schemas as motive reconsidered. In *Human motives and cultural models*, edited by R. G. D'Andrade and C. Strauss, 191–224. Cambridge: Cambridge University Press.
- Strauss, C., and N. Quinn. 1997. *A cognitive theory of cultural meaning*. New York: Cambridge University Press.
- Streib, G. F. 1952. Use of survey methods among the Navaho. *American Anthropologist* 54:30–40.

- Stricker, L. J. 1988. Measuring social status with occupational information: A simple method. *Journal of Applied Social Psychology* 18:423–37.
- Stunkard, A., and D. Kaplan. 1977. Eating in public places: A review of reports of the direct observation of eating behavior. *International Journal of Obesity* 1:89–101.
- Sturrock, K., and J. Rocha. 2000. A multidimensional scaling stress evaluation table. *Field Methods* 12:49–60.
- Sturtevant, W. C. 1959. A technique for ethnographic note-taking. *American Anthropologist* 61:677–78.
- Sudman, S. 1976. *Applied sampling*. New York: Academic Press.
- Sudman, S., and N. M. Bradburn. 1974. *Response effects in surveys: Review and synthesis*. Chicago: Aldine.
- Sudman, S., and N. M. Bradburn. 1982. *Asking questions*. San Francisco: Jossey-Bass.
- Sudman, S., N. M. Bradburn, and N. Schwarz. 1996. *Thinking about answers: The application of cognitive processes to survey methodology*. San Francisco: Jossey-Bass.
- Sudman, S., and G. Kalton. 1986. New developments in the sampling of special populations. *Annual Review of Sociology* 12:401–29.
- Sudman, S., and N. Schwarz 1989. Contributions of cognitive psychology to advertising research. *Journal of Advertising Research* 29:43–53.
- Sugarman, J. R., G. Brennehan, W. LaRoque, C. W. Warren, and H. I. Goldberg. 1994. The urban Indian oversample in the 1988 National Maternal and Infant Health Survey. *Public Health Reports* 109:243–50.
- Sugihara, Y., and E. Katsurada. 2000. Gender-role personality traits in Japanese culture. *Psychology of Women Quarterly* 24:309–18.
- Sugihara, Y., and J. A. Warner 1999. Endorsements by Mexican-Americans of the Bem Sex-Role Inventory: Cross-ethnic comparison. *Psychological Reports* 85:201–11.
- Sugita, E. 2004. Domestic water use, hygiene behavior, and children's diarrhea in rural Uganda. Ph.D. diss., University of Florida.
- Sullivan, K. A. 1990. Daily time allocation among adult and immature yellow-eyed juncos over the breeding season. *Animal Behaviour* 39:380–88.
- Sullivan, P., and K. Elifson. 1996. In the field with snake handlers. In *In the field. Readings on the research experience*, edited by C. D. Smith and W. Kornblum, 33–38. New York: Praeger.
- Sundvik, L., and M. Lindeman. 1993. Sex-role identity and discrimination against same-sex employees. *Journal of Occupational and Organizational Psychology* 66:1–11.
- Swan, D. C., and G. R. Campbell. 1989. Differential reproduction rates and Osage population change, 1877–1907. *Plains Anthropologist* 34(124, pt. 2): 61–74.
- Swora, M. G. 2003. Using cultural consensus analysis to study sexual risk perception: A report on a pilot study. *Culture, Health, and Sexuality* 5:339–52.
- Sykes, R. E., and E. E. Brent. 1983. *Policing: A social behaviorist perspective*. New Brunswick, N.J.: Rutgers University Press.
- Sykes, R. E., R. D. Rowley, and J. M. Schaefer. 1990. Effects of group participation on drinking behaviors in public bars: An observational survey. *Journal of Social Behavior and Personality* 5(4): 385–402 (special issue).
- Synodinos, N. E., and S. Yamada. 2000. Response rate trends in Japanese surveys. *International Journal of Public Opinion Research* 12(1): 48–72.
- Szalai, A., ed. 1972. *The use of time. Daily activities of urban and suburban populations in twelve countries*. The Hague: Mouton.
- Tannen, D. 1984. *Conversational style: Analyzing talk among friends*. Norwood, N.J.: Ablex.
- Tashakkori, A. and C. Teddlie. 2003. *Handbook of mixed methods in social and behavioral research*. Thousand Oaks, Calif.: Sage.

- Taylor, H. 1997. The very different methods used to conduct telephone surveys of the public. *Journal of the Marketing Research Society* 37:421–32.
- Taylor, S. J. 1991. Leaving the field: Relationships and responsibilities. In *Experiencing fieldwork: An inside view of qualitative research*, edited by W. B. Shaffir and R. A. Stebbins, 238–45. Newbury Park, Calif.: Sage.
- Taylor, S. J., and R. Bogdan. 1984. *Introduction to qualitative research methods*. 2d ed. New York: Wiley.
- Tedlock, D. 1977. Toward an oral poetics. *New Literary History* 8:507–19.
- Tedlock, D. 1987. Hearing a voice in an ancient text: Quiché Maya poetics in performance. In *Native American discourse: Poetics and rhetoric*, edited by J. Sherzer and A. Woodbury, 140–75. Cambridge: Cambridge University Press.
- Tesch, R. 1990. *Qualitative research: Analysis types and software tools*. New York: Falmer Press.
- Tewksbury, R. 2002. Bathhouse intercourse: Structural and behavioral aspects of an erotic oasis. *Deviant Behavior* 23:75–112.
- Thakur, H. B. 1996. Practical steps towards saving the lives of 25,000 potential victims of dowry and bride-burning in India in the next four years. *Journal of South Asia Women Studies* 2(2): <http://www1.shore.net/~india/jsaws/issue3/>.
- Thomas, D., E. Frankenber, and J. P. Smith. 2001. Lost but not forgotten: Attrition and follow-up in the Indonesia family life survey. *The Journal of Human Resources* 36:556–92.
- Thomas, J. S. 1981. The economic determinants on leadership on a Tojalabal Maya community. *American Ethnologist* 8:127–38.
- Thurman, Q., S. Jackson, and J. Zhao. 1993. Drunk-driving research and innovation: A factorial survey study of decisions to drink and drive. *Social Science Research* 22:245–64.
- Tinsley, H. E., D. J. Tinsley, and C. E. Croskeys. 2002. Park usage, social milieu, and psychosocial benefits of park use reported by older urban park users from four ethnic groups. *Leisure Sciences* 24:199–218.
- Tooker, E. 1997. Introduction. In *Systems of consanguinity and affinity of the human family*, edited by L. H. Morgan. Lincoln: University of Nebraska Press.
- Torgerson, W. S. 1958. *Theory and methods of scaling*. New York: Wiley.
- Toulmin, S. E. 1980. Philosophy of science. *Encyclopaedia britannica*, Vol. 16. Chicago: Encyclopaedia Britannica.
- Tourangeau, R., L. J. Rips, and K. Rasinski. 2000. *The psychology of survey response*. New York: Cambridge University Press.
- Tourangeau, R., and T. W. Smith. 1996. Asking sensitive questions: The impact of data collection, question format, and question context. *Public Opinion Quarterly* 60:275–304.
- Tremblay, M. 1957. The key informant technique: A non-ethnographical application. *American Anthropologist* 59:688–701.
- Trent, R., and H. R. Bernard. 1985. Local support for an innovative transit system. *Journal of Advanced Transportation Research* 19:237–39.
- Trost, J. E. 1986. Statistically nonrepresentative stratified sampling: A sampling technique for qualitative studies. *Qualitative Sociology* 9:54–57.
- Trotter, R. T., II. 1981. Remedios caseros: Mexican-American home remedies and community health problems. *Social Science and Medicine* 15B:107–14.
- Truex, G. 1993. Tagging and tying: Notes on codes in anthropology. *Cultural Anthropology Methods Journal* 5(1): 3–5.
- Tudor-Locke, C., K. A. Lind, J. P. Reis, B. E. Ainsworth, and C. A. Macera. 2004. A preliminary evaluation of a pedometer-assessed physical activity self-monitoring survey. *Field Methods* 16:422–38.
- Tuladhar, S. M., S. S. Acharya, M. Pradhan, J. Pollock, and G. Dallabetta. 1998. The role of

- pharmacists in HIV/STD prevention: Evaluation of an STD syndromic management intervention in Nepal. *AIDS* 12(Suppl. 2): S81–S87.
- Turley, S. K. 1999. A case of response rate success. *Journal of the Market Research Society* 41(3): 301–309.
- Turnbull, C. 1972. *The mountain people*. New York: Simon & Schuster.
- Turnbull, C. 1986. Sex and gender: The role of subjectivity in field research. In *Self, sex and gender in cross-cultural fieldwork*, edited by T. L. Whitehead and M. E. Conaway, 17–29. Urbana: University of Illinois Press.
- Twenge, J. M. 1997. Attitudes towards women, 1970–1995. *Psychology of Women Quarterly* 21:35–51.
- Twisk, J., and W. de Vente. 2002. Attrition in longitudinal studies: How to deal with missing data. *Journal of Clinical Epidemiology* 55:329–37.
- Tyler, S. A., ed. 1969. *Cognitive anthropology*. New York: Holt, Rinehart and Winston.
- Uchendu, V. 1965. *The Igbo of Southeastern Nigeria*. New York: Holt, Rinehart & Winston.
- Umezaki, M., T. Yamauchi, and R. Ohtsuka. 2002. Time allocation to subsistence activities among the Huli in rural and urban Papua New Guinea. *Journal of Biosocial Science* 34:133–37.
- UNESCO 1989a. *Thematic list of descriptors—Anthropology*. London: Routledge.
- UNESCO 1989b. *Thematic list of descriptors—Political Science*. London: Routledge.
- UNESCO 1989c. *Thematic list of descriptors—Sociology*. London: Routledge
- University of Michigan. Survey Research Center. 1976. *Interviewer's Manual*. Rev. ed. Ann Arbor: Institute for Social Research, University of Michigan.
- U.S. Bureau of the Census n.d. Statement by William G. Barron Jr. on the current status of results of Census 2000 Accuracy and Coverage Evaluation Survey. CB01-CS.06. <http://www.census.gov/Press-Release/www/releases/archives/census2000/000710.html>.
- Vadez, V., V. Reyes-García, R. Godoy, L. Williams, L. Apaza, E. Byron, T. Huanca, W. R. Leonard, E. Pérez, and D. Wilkie. 2003. Validity of self-reports to measure deforestation: Evidence from the Bolivian highlands. *Field Methods* 15:289–304.
- van der Geest, S., and S. Sarkodie. 1998. The fake patient: A research experiment in a Ghanaian hospital. *Social Science & Medicine* 47(9): 1373–81.
- Van Hattum, M.J.C., and E. D. de Leeuw. 1999. A disk-by-mail survey of pupils in primary schools: Data quality and logistics. *Journal of Official Statistics* 15:413–29.
- Van Maanen, J. 1983. *Qualitative methodology*. Beverly Hills, Calif.: Sage.
- Van Maanen, J., M. Miller, and J. C. Johnson. 1982. An occupation in transition: Traditional and modern forms of commercial fishing. *Work and Occupations* 9:193–216.
- Van Willigen, J., and V. C. Channa. 1991. Law, custom, and crimes against women: The problem of dowry death in India. *Human Organization* 50:369–77.
- Vannatta, R. A. 1996. Risk factors related to suicidal behavior among male and female adolescents. *Journal of Youth and Adolescence* 25:149–60.
- Vasquez, L. A., E. Garcia-Vasquez, S. A. Bauman, and A. S. Sierra. 1997. Skin color, acculturation, and community interest among Mexican American students: A research note. *Hispanic Journal of Behavioral Sciences* 19:377–86.
- Vaughn, S., J. S. Schumm, and J. M. Sinagub. 1996. *Focus group interviews in education and psychology*. Thousand Oaks, Calif.: Sage.
- Veatch, H. B. 1969. *Two logics: The conflict between classical and neoanalytic philosophy*. Evanston, Ill.: Northwestern University Press.
- Vieth, A. Z., K. K. Hagglund, D. L. Clay, and R. G. Frank. 1997. The contribution of hope and affectivity to diabetes-related disability: An exploratory study. *Journal of Clinical Psychology in Medical Settings* 4:65–77.

- Vincke, J., R. Bolton, and P. De Vleeschouwer. 2001. The cognitive structure of the domain of safe and unsafe gay sexual behavior in Belgium. *AIDS CARE* 13:57–70.
- Voltaire. 1967 [1738]. *The elements of Sir Isaac Newton's philosophy*. Translated by J. Hanna. London: Cass.
- Wagley, C. 1983. Learning fieldwork: Guatemala. In *Fieldwork: The human experience*, edited by R. Lawless, V. H. Sutlive, and M. D. Zamora, 1–18. New York: Gordon & Breach.
- Wagner, P. 2001. *A history and theory of the social sciences: Not all that is solid melts into air*. Thousand Oaks, Calif.: Sage.
- Wainer, H. 1999. One cheer for null hypothesis significance testing. *Psychological Methods* 4(2): 212–13.
- Waitzkin, H., T. Britt, and C. Williams. 1994. Narratives of aging and social problems in medical encounters with older persons. *Journal of Health and Social Behavior* 35:322–48.
- Wakin, E. 1992. *Anthropology goes to war: Professional ethics and counterinsurgency in Thailand*. Madison: University of Wisconsin, Center for Southeast Asian Studies.
- Wall Street Journal, The*. 1986. 4 September, p. 29.
- Wallace, A.F.C. 1962. Culture and cognition. *Science* 135:351–57.
- Wallace, A.F.C. 1965. Driving to work. In *Context and meaning in cultural anthropology*, edited by M. E. Spiro, 277–96. New York: The Free Press.
- Wallerstein, E. M. 1974. *The modern world-system: Capitalist agriculture and the origins of the European world-economy*. New York: Academic Press.
- Wallerstein, E. M. 2004. *World-systems analysis: An introduction*. Durham, N.C.: Duke University Press.
- Waltermaurer, E. M., C. A. Ortega, and L.-A. McNutt. 2003. Issues in estimating the prevalence of intimate partner violence: Assessing the impact of abuse status on participation bias. *Journal of Interpersonal Violence* 18:959–74.
- Wanous, J. P., A. E. Reichers, and M. J. Hudy. 1997. Overall job satisfaction: How good are single-item measures? *Journal of Applied Psychology* 82:247–52.
- Ward, V. M., J. T. Bertrand, and L. F. Brown. 1991. The comparability of focus group and survey results. *Evaluation Review* 15:266–83.
- Warner, W. L., ed. 1963. *Yankee City*. New Haven, Conn.: Yale University Press.
- Warner, W. L., and P. S. Hunt. 1941. *The social life of a modern community*. New Haven, Conn.: Yale University Press.
- Warren, C.A.B. 1988. *Gender issues in field research*. Newbury Park, Calif.: Sage.
- Warriner, K., J. Goyder, H. Gjertsen, P. Hohner, and K. McSpurren. 1996. Charities, no; lotteries, no; cash, yes. Main effects and interactions in a Canadian incentives experiment. *Public Opinion Quarterly* 60:542–62.
- Warwick, D. P., and C. A. Lininger. 1975. *The sample survey: Theory and practice*. New York: McGraw-Hill.
- Watson, O. M., and T. D. Graves. 1966. Quantitative research in proxemic behavior. *American Anthropologist* 68:971–85.
- Wax, R. 1971. *Doing fieldwork: Warnings and advice*. Chicago: University of Chicago Press.
- Wax, R. 1986. Gender and age in fieldwork and fieldwork education: "Not any good thing is done by one man alone." In *Self, sex and gender in cross-cultural fieldwork*, edited by T. L. Whitehead and M. E. Conaway, 129–50. Urbana: University of Illinois Press.
- Weatherford, J. M. 1986. *Porn row*. New York: Arbor House.
- Webb, B. 1926. *My apprenticeship*. London: Longmans, Green.
- Webb, E. J., and K. E. Weick. 1983. Unobtrusive measures in operational theory: A reminder. In *Qualitative methodology*, edited by J. Van Maanen, 209–24. Newbury Park, Calif.: Sage.

- Webb, E. J., D. T. Campbell, R. D. Schwartz, and L. Sechrest. 1966. *Unobtrusive measures: Non-reactive research in the social sciences*. Chicago: Rand McNally.
- Webb, S., and B. P. Webb. 1910. *The state and the doctor*. New York: Longmans, Green.
- Weeks, M. F., and R. P. Moore. 1981. Ethnicity-of-interviewer effects on ethnic respondents. *Public Opinion Quarterly* 45:245–49.
- Weinberger, M. 1973. Getting the quota sample right. *Journal of Advertising Research* 13(5): 69–72.
- Weisner, T. 2002. The American dependency conflict: Continuities and discontinuities in behavior and values of countercultural parents and their children. *Ethos* 29:271–95.
- Weitzman, E. A., and M. Miles. 1995. *Computer programs for qualitative data analysis: A software sourcebook*. Thousand Oaks, Calif.: Sage.
- Weller, S. C. 1983. New data on intracultural variability: The hot-cold concept of medicine and illness. *Human Organization* 42:249–57.
- Weller, S. C. 2004. Measuring cultural beliefs: The cultural consensus model. In *Encyclopedia of social measurement*, edited by K. Kempf-Leonard. Vol. 1, A–E, 579–85. Oxford: Elsevier/Academic Press.
- Weller, S. C., and C. I. Dungy. 1986. Personal preferences and ethnic variations among Anglo and Hispanic breast and bottle feeders. *Social Science and Medicine* 23:539–48.
- Weller, S. C., and A. K. Romney. 1988. *Structured interviewing*. Newbury Park, Calif.: Sage.
- Wells, G. L., E. A. Olson, and S. D. Charman. 2003. Distorted retrospective eyewitness reports as functions of feedback and delay. *Journal of Experimental Psychology: Applied* 9:42–52.
- Wentland, E. J., and K. W. Smith. 1993. *Survey responses: An evaluation of their validity*. San Diego: Academic Press.
- Weppner, R. S. 1973. An anthropological view of the street addict's world. *Human Organization* 32:111–21.
- Weppner, R. S., ed. 1977. *Street ethnography: Selected studies of crime and drug use in natural settings*. Beverly Hills, Calif.: Sage.
- Werblow, J. A., H. M. Fox, and A. Henneman. 1978. Nutrition knowledge, attitudes and patterns of women athletes. *Journal of the American Dietetic Association* 73:242–45.
- Werner, D. 1985. Psycho-social stress and the construction of a flood-control dam in Santa Catarina, Brazil. *Human Organization* 44:161–66.
- Werner, O. 1994. Ethnography and translation. Issues and challenges. *Sartonian* 7:59–135.
- Werner, O. 2000a. How to reduce an unwritten language to writing: I (vowels). *Field Methods* 12:61–71.
- Werner, O. 2000b. How to reduce an unwritten language to writing: II (consonants). *Field Methods* 12:239–50.
- Werner, O. 2001. How to reduce an unwritten language to writing: III (phonetic similarity, suspicious pairs, and minimal pairs). *Field Methods* 13:97–102.
- Werner, O. 2002a. How to reduce an unwritten language to writing: IV. Complementary distribution. *Field Methods* 14:217–27.
- Werner, O. 2002b. How to reduce an unwritten language to writing: V. Problems with phonemes. *Field Methods* 14:337–42.
- Werner, O., and D. T. Campbell. 1970. Translating, working through interpreters, and the problem of decentering. In *Handbook of method in cultural anthropology*, edited by R. Naroll and R. Cohen, 398–420. New York: Natural History Press.
- Werner, O., and J. Fenton. 1973. Method and theory in ethnoscience or ethnoepistemology. In *A handbook of method in cultural anthropology*, edited by R. Naroll and R. Cohen, 537–78. New York: Columbia University Press.
- Werner, O., and G. M. Schoepfle. 1987. *Systematic fieldwork*. 2 vols. Newbury Park, Calif.: Sage.

- Wertz, R. T. 1987. Language treatment for aphasia is efficacious, but for whom? *Topics in Language Disorders* 8:1–10.
- West, S. G., S. P. Gunn, and P. Chernicky. 1975. Ubiquitous Watergate: An attributional analysis. *Journal of Personality and Social Psychology* 32:55–65.
- White, K. R. 1980. Socio-economic status and academic achievement. *Evaluation in Education* 4:79–81.
- White, L. A. 1959. *The evolution of culture. The development of civilization to the fall of Rome*. New York: McGraw-Hill.
- White, L. A. 1963. *The ethnography and ethnology of Franz Boas*. Austin: The Museum of the University of Texas.
- Whitehead, T. L., and M. E. Conaway, eds. 1986. *Self, sex and gender in cross-cultural fieldwork*. Urbana: University of Illinois Press.
- Whiting, B. W., and J.W.M. Whiting. 1973. Methods for observing and recording behavior. In *Handbook of method in cultural anthropology*, edited by R. Naroll and R. Cohen, 282–315. New York: Columbia University Press.
- Whiting, B. W., and J.W.M. Whiting (with R. Longabaugh). 1975. *Children of six cultures: A psycho-cultural analysis*. Cambridge: Harvard University Press.
- Whiting, J.W.M. 1982. Standards for psychocultural research. In *Crisis in anthropology. View from Spring Hill, 1980*, edited by E. A. Hoebel, R. Currier, and S. Kaiser, 155–64. New York: Garland.
- Whiting, J.W.M., I. L. Child, and W. W. Lambert et al. 1966. *Field guide for a study of socialization*. New York: Wiley.
- WHO. 1993. *Focused ethnographic study of acute respiratory infections*. Division for the Control of Diarrhoeal and Respiratory Diseases, ARI/93.2. Geneva: WHO.
- Whyte, W. F. 1960. Interviewing in field research. In *Human organization research*, edited by R. W. Adams and J. J. Preiss, 299–314. Homewood, Ill.: Dorsey.
- Whyte, W. F. 1981 [1943]. *Street corner society: The social structure of an Italian slum*. 3rd ed. Chicago: University of Chicago.
- Whyte, W. F. 1984. *Learning from the field: A guide from experience*. Newbury Park, Calif.: Sage.
- Whyte, W. F. 1989. Doing research in Cornerville. In *In the field: Readings on the field research experience*, edited by C. D. Smith and W. Kornblum, 69–82. New York: Praeger.
- Whyte, W. F. 1996a. Qualitative sociology and deconstructionism. *Qualitative Inquiry* 2:220–26.
- Whyte, W. F. 1996b. Facts, interpretations and ethics in qualitative inquiry. *Qualitative Inquiry* 2:242–44.
- Whyte, W. F., and K. K. Whyte 1984. *Learning from the field: A guide from experience*. Newbury Park, Calif.: Sage.
- Wiederman, M. 1993. Evolved gender differences in mate preferences: Evidence from personal advertisements. *Ethology and Sociobiology* 14(5): 331–51.
- Wiederman, M., D. Weis, and E. Algeier. 1994. The effect of question preface on response rates in a telephone survey of sexual experience. *Archives of Sexual Behavior* 23:203–15.
- Wilke, J. R. 1992. Supercomputers manage holiday stock. *The Wall Street Journal*, 23 December, B1:8.
- Willett, W. C., M. J. Stampfer, J. E. Manson, G. A. Colditz, B. A. Rosner, F. E. Speizer, and C. H. Hennekens. 1996. Coffee consumption and coronary heart disease in women: A ten-year follow-up. *The Journal of the American Medical Association* 275:458–62.
- Williams, B. 1978. *A sampler on sampling*. New York: Wiley.
- Williams, H. A. 1995. Social support, social networks and coping of parents of children with cancer: Comparing White and African American parents. Ph.D. diss., University of Florida.

- Williams, T. 1996. Exploring the cocaine culture. In *In the field. Readings on the research experience*, edited by C. D. Smith and W. Kornblum, 27–32. New York: Praeger.
- Willimack, D. K., H. Schuman, B-E. Pennell, and J. M. Lepkowski. 1995. Effects of a prepaid nonmonetary incentive on response rates and response quality in face-to-face survey. *Public Opinion Quarterly* 59:78–92.
- Willis, F. N., and R. A. Carlson. 1993. Singles ads: Gender, social class, and time. *Sex Roles* 29(5–6): 387–404.
- Willms, D. G., J. A. Best, D. W. Taylor, J. R. Gilbert, D.M.C. Wilson, E. A. Lindsay, and J. Singer. 1990. A systematic approach for using qualitative methods in primary prevention research. *Medical Anthropology Quarterly* 4:391–409.
- Wilson, D., J. McMaster, R. Greenspan, L. Mboyi, T. Ncube, and B. Sibanda. 1990. Cross-cultural validation of the Bem Sex Role Inventory in Zimbabwe. *Personality and Individual Differences* 11(7): 651–66.
- Wilson, R. P., M. Nxumalo, B. Magonga, Jr., G. A. Shelley, K. A. Parker, and Q. Q. Dlamini. 1993. Diagnosis and management of acute respiratory infections by Swazi child caretakers, healers, and health providers, 1990–1991. Office of Analysis and Technical Reports, United States Agency for International Development. Working Paper.
- Windelband, W. 1998 [1894]. History and natural science. *Theory and Psychology* 8:5–22.
- Wittfogel, K. A. 1957. *Oriental despotism; a comparative study of total power*. New Haven, Conn.: Yale University Press.
- Wolcott, H. 1995. *The art of fieldwork*. Walnut Creek, Calif.: AltaMira.
- Wolcott, H. 1999. *Ethnography: A way of seeing*. Walnut Creek, Calif.: AltaMira.
- Wolcott, H. 2002. *Sneak kid and its aftermath: Ethics and intimacy in fieldwork*. Walnut Creek, Calif.: AltaMira.
- Wolcott, H. 2005. *The art of fieldwork*. 2d ed. Walnut Creek, Calif.: AltaMira.
- Wolf, D. R. 1991. High-risk methodology. Reflections on leaving an outlaw society. In *Experiencing fieldwork*, edited by W. B. Shaffir and R. A. Stebbins, 211–23. Newbury Park, Calif.: Sage.
- Woods, P. 1986. *Inside schools; ethnography in educational research*. New York: Routledge and Kegan Paul.
- Woolhouse, R. S. 1996. Locke. In *The Blackwell companion to philosophy*, edited by N. Bunnin and E. P. Tsui-James, 541–54. Oxford: Blackwell.
- Worth, S., and J. Adair. 1970. Navajo filmmakers. *American Anthropologist* 72:9–33.
- Worth, S., and J. Adair. 1972. *Through Navajo eyes: An exploration in film communication and anthropology*. Albuquerque: University of New Mexico Press.
- Wuthnow, R. 1976. A longitudinal, cross-national indicator of societal religious commitment. *Journal for the Scientific Study of Religion* 16:87–99.
- Yammarino, F. J., S. J. Skinner, and T. L. Childers. 1991. Understanding mail survey response behavior. A meta-analysis. *Public Opinion Quarterly* 55:613–39.
- Yancey, G. A., and S. W. Yancey. 1997. Black-white differences in the use of personal advertisements for individuals seeking interracial relationships. *Journal of Black Studies* 27:650–67.
- Yoder, P. S., R. Drew, and Z. Zhong. n.d. Knowledge and practices related to diarrhea disorders, oral rehydration therapy, and vaccinations in Lubumbashi, Zaire. Philadelphia: University of Pennsylvania HEALTHCOM Project, pp. 12–14.
- Young, J. C. 1978. Illness categories and action strategies in a Tarascan town. *American Ethnologist* 5:81–97.
- Young, J. C. 1980. A model of illness treatment decisions in a Tarascan town. *American Ethnologist* 7:106–31.

- Young, J. C., and L. Y. Garro. 1982. Variation in the choice of treatment in two Mexican communities. *Social Science and Medicine* 16:1453–63.
- Young, J. C., and L. C. Garro. 1994 [1981]. *Medical choice in a Mexican village*. Prospect Heights, Ill.: Waveland.
- Yu, X. 1995. Conflict in a multicultural organization: An ethnographic attempt to discover work-related cultural assumptions between Chinese and American co-workers. *International Journal of Conflict Management* 6:211–32.
- Zacharias, J., T. Stathopoulos, and H. Wu. 2004. Spatial behavior in San Francisco's plazas: The effects of microclimate, other people, and environmental design. *Environment and Behavior* 36:638–58.
- Zahavi, D. 2003. *Husserl's phenomenology*. Stanford, Calif.: Stanford University Press.
- Zehner, R. B. 1970. Sex effects in the interviewing of young adults. *Sociological Focus* 3:75–84.
- Zhang, J., J. M. Norvilitis, and S. Jin. 2001. Measuring gender orientation with the Bem Sex Role Inventory in Chinese culture. *Sex Roles* 44:237–51.
- Zinner, D., J. Hindahl, and M. Schwibbe. 1997. Effects of temporal sampling patterns of all-occurrence recording in behavioural studies: Many short sampling periods are better than a few long ones. *Ethology* 103(3): 236–46.
- Zipp, J. F., and J. Toth. 2002. She said, he said, they said: The impact of spousal presence in survey research. *Public Opinion Quarterly* 66:177–208.
- Znaniecki, F. 1963 [1952]. *Cultural sciences. Their origin and development*. Urbana: University of Illinois Press.
- Zorbaugh, H. W. 1929. *The Gold Coast and the slum. A sociological study of Chicago's near North Side*. Chicago: The University of Chicago Press.

Subject Index

Numbers in italics refer to figures and tables.

- AAA (American Anthropological Association), 74, 279
- abscissa, 572, 603
- absolute value, 598
- abstracts, 101
- Abstracts in Anthropology (AIA), 101
- A-CASI (audio, computer-assisted, self-administered interview), 253, 270
- acculturation factor, 200
- accuracy, 54–55, 245–50; jogging informants' memories, 247–50; self-reported behavior, 246–47; of samples, 166; social desirability effect, 250
- acquiescence effect, 241–42, 259
- adjectives of activity, 338
- adjectives of evaluation, 338
- adjectives of potency, 338
- age cohorts, 154
- aggregate proximity matrix, 687–88
- A&HCI (*Arts and Humanities Citation Index*), 98, 99
- AIA (Abstracts in Anthropology), 101
- aided recall, 249
- AIO (*Anthropological Index Online*), 100
- AL (*Anthropological Literature*), 100–101
- algorithm, 461, 462
- alphabetic cues, 248, 302
- alpha level, 585
- alternative hypothesis, 584
- American Anthropological Association (AAA), 74, 279
- American marriage schema, 482–83
- amount of control, 210
- analysis of covariance (ANCOVA), 602–3
- analysis of variance (ANOVA), 598–603; calculation, 598–601; case examples, 601–2; testing the significance of, 601; uses, 602–3
- analytic categories, 492
- analytic induction, 522, 544–48
- analytic notes, 398
- ANCOVA (analysis of covariance), 602–3
- Annual Review* series, 97, 98
- ANOVA. *See* analysis of variance
- AnSWR (software), 521
- antecedent variables, 63, 607
- Anthropac* (software), 206, 301, 307, 310–11, 326, 547
- Anthropological Index Online* (AIO), 100
- Anthropological Literature* (AL), 100–101
- Anthropology Newsletter*, 279
- Anthropology Plus*, 101
- archival research, 448–50
- Aristotle, 6
- Arouet, François Marie, 11
- arrays, 688
- artifacts, 90, 463
- Arts and Humanities Citation Index* (A&HCI), 98, 99

- association. *See* covariation
Atlas/ti (software), 521
Attitude Toward Women Scale (AWS), 44–45
attrition, 291
average. *See* mean
AWS (Attitude Toward Women Scale), 44–45
- Babbage, Charles, 414
back translation, 277–78
Bacon, Francis, 9
Badji, Antoine, 278
baiting, 221–22
balanced incomplete block design (BIB), 309–11, 310
Barbeau, Marius, 465
bar graphs, 572, 573
Barnett, Steven, 424
Bartlett, F. C., 109
baseline data, 133–35
Bastian, Adolph, 464
Bateson, Gregory, 464
batteries, 259, 259
b coefficients, 663
becoming the phenomenon, 371
B–E convention, 407–8
behavior, 90–91, 463; inaccuracy of self-reported, 246–47
behavior trace studies, 444–48; pros and cons, 447–48
BEM Sex Role Inventory (BSRI), 32–33
Bernoulli, Daniel, 16
best fitting line. *See* regression line
between-group variance, 163, 598, 599, 600; maximizing, 154, 162–66
Beynon, William, 465–66
BIB (balanced incomplete block design), 309–11, 310
biblical exegesis, 473–74
bibliographic tools. *See* online databases
biconditional statement, 547
bimodal distribution, 171, 172, 575, 579, 579–80, 580
- Bismarck, Otto von, 16
bivariate analysis, 549–50, 594–637; analysis of variance, 598–603; chi-square test, 609–12, 621; covariation and, 603–7; cross-tabulation tables, 605, 605, 606; eta-squared calculation, 621, 635, 637; Fisher’s exact test, 614–15; gamma, 615–19, 621; Kendall’s T_b , 619–20; nonlinear relations, 632–35; odds ratio, 613; Pearson’s product moment correlation (*r*), 621, 622–24, 626, 629–30, 630–32; proportionate reduction of error (PRE), 608–9; regression, 625–29; r-squared calculation, 623, 629–30; Spearman’s *r*, 621–22; *t*-test, 595–98; Yule’s *Q*, 620–21
bivariate correlations, 95
bivariate tables, 605, 605, 606; conventions, 605–7; expected frequencies for chi-squared, 610–12
Black Elk (Oglala Sioux), 472
Black Elk Speaks (Neihardt), 472
Black Like Me (Griffin), 437
Blowsnake, Sam, 466
Blumberg, Albert, 19
Boas, Franz, 26, 82, 199, 360, 391–92, 464–65, 470, 475
Bohannon, Paul, 469
Bonferroni correction, 640–41
Bonilla, Zobeida, 190
The Book of Chilam Balam of Chumayel, 478
Boolean logic, 545–47
Boolean operators, 107
box-and-whisker plots, 576–77, 577
box plots, 576–77, 577
Boys in White (Becker), 346
branching tree diagrams, 539, 540
Brandes, Stanley, 290
Brim, John, 311
BSRI (BEM Sex Role Inventory), 32–33
Burgess, Ernest, 346
- Campbell and Stanley posttest-only design, 123

- Campos, Alberto, 477
Cantril ladder of life, 339, 339–40
CAPI (computer-assisted personal interviewing), 254, 270
case control design, 187
case control sampling, 194–95
CASI (computer-assisted self-administered interview), 258–59, 270
cassette tape recorders, 229–30
CATI (computer-assisted telephone interviewing), 252–53, 263
causal analysis, 541
causal flow charts, 460–62
causal maps, 460–61
causal modeling, 649
cause and effect, 28, 60–66
Census, U.S., 147
central limit theorem, 173–78
central tendency, measures of, 562–72;
 mean, 562, 566–72; median, 562,
 563–66; mode, 562–63
chain referral sampling, 192–94
character variables, 555
Chicago School, 346
Chicago Tribune, 188
Chinookan, 475
chi-square test: bivariate analysis, 609–
 12, 621; testing significance of 592;
 univariate analysis, 591–92, 593
Chomsky, N., 4
C-I-SAID (software), 521
CIS (Congressional Information Service),
 104–5
Clark, Sam, 290
classic experiment, 110–12
closed-ended questions, 268–70
closest-neighbor analysis, 690–91, 692
clumps, 688
cluster analysis, 689–92, 690, 691, 692
clusters, 688
cluster sampling, 157–58
codebooks, 550, 553; building of, 552,
 554–55
code notes, 499–500
Codex Borgia, 508
Codex Nuttall, 508
coding, 403; coding film 403; coding vs.
 indexing, 405; content analysis and,
 510–15; continuous monitoring,
 419–25; cross-cultural content analysis
 and, 516–18; data cleaning, 555; field
 notes and, 398–405, 399, 402, 407;
 grounded theory text analysis and,
 492–96, 497, 498–99; intercoder reli-
 ability, 511, 512–15; mechanics of,
 404–5, 405; published codes, 518;
 time allocation studies and, 431–32;
 univariate analysis and, 550–55
coefficient of determination, 629
coefficient of reproducibility (CR),
 323–24
cognitive dissonance theory, 64, 65, 130
cognitive testing, 286–87
Cohen's Kappa, 513–14
comparative field experiments, 138–40
complete-link clustering, 690, 691, 692
complete observer, 347
complete participant, 347
complex sampling designs, 158
complex scales, 319–20
componential analysis, 522, 541–44, 542,
 543
composite measures, 318–41; Cantril
 ladder of life, 339, 339–40; faces
 scale, 340, 340–41; Guttman scale,
 321–27, 322, 325; indexes, 320–21;
 Likert scales, 327–35; multiple indica-
 tor scales, 319–20; semantic differen-
 tial scaling, 274, 336–38, 337; single
 indicator scales, 318–19, 320; types of,
 318
computers: questionnaires and, 252–55;
 text analysis and, 519–21. *See also*
 specific software
Comte, Auguste, 11, 12, 13, 14
concepts, 48
conceptual definitions, 37–40
conceptual model, 657, 658

- concordances, 495
 conditional statements, 547
 condition index, 672
 conditions, 144
 Condon, Richard, 377
 Condorcet, Marie Jean de, 11
 confidence intervals, 178–79, 184; small samples and, 180
 confidence limits, 589
 confirmatory research, 110
 confirmatory stage, 493
 confounds, 113, 114–18
 Congressional Information Service (CIS), 104–5
 constructivist view, 3
 construct validity, 58
 consumer behavior, monitoring of, 26
 Consumer Price Index, U.S., 320
 contagion theory, 65
Contemporary Ethnography, 346
 content analysis, 473, 505, 507–15; coding, 510–15; sampling, 509–10. *See also* text analysis
 content analysis, cross-cultural, 515–19; coding, 516–18; other problems, 518–19; sampling, 516
 content validity, 57–58
 context units, 510
 contingencies, 272
 contingency flow chart, 272, 273
 continuous monitoring (CM), 413–25; case examples, 415–19; coding, 419–25; ethograms, 414; reactivity and, 424–25; time allocation studies, mixing with, 434–35; uses, 413–14; video and, 423–25. *See also* direct observation
 control, 62
 control case, 195
 control group, 110, 111
 controlling for, 649
 convenience sampling, 187, 191–92
 conversation analysis, 486–91; adjacency pairs, 490–91; taking turns, 487–90; transcriptions, 487, 488. *See also* discourse analysis
 correlation. *See* covariation
The Counter-Revolution of Science (Hayek), 20
The Course of Positive Philosophy (Comte), 13
 covariation, 61–62, 94–95, 594; bivariate analysis and, 603–7; direction and shape, 603–5
 covert categories, 534, 535, 536, 539
Crashing Thunder (Blowsnake), 466
 CR (coefficient of reproducibility), 323–24
 criterion referenced, 320
 criterion validity, 58–59
 critical case studies, 190, 191
 critical region, 585
 critical value of t , 181
 Cronbach's alpha, 331, 332–34
Cross-Cultural Research, 518
 cross-sectional surveys, 287–88
 cross-tabulation tables, 605, 605, 606
 cued recall, 248–49
 cultural consensus model, 203–7, 204–5
 cultural consonance, 65–66
 cultural data, 146–47, 187
 cultural domain analysis, 299–317; free listing, 301–5; free pile sorts, 311–15, 315; paired comparisons, 315–16, 316; rank ordering, 316–17; rating scales, 317; sentence frame technique, 305–7; triad tests, 307–11. *See also* structured interviewing
 cultural domains, folk taxonomies and, 536–37
 cultural paradigm, 79
 cultural relativism, 4–5
 culture shock, 379–81
 cumulative index, 320–21
 cumulative knowledge, 113
 cumulative mean, 173
 D'Alambert, Jean, 16
 Daniels, Arlene Kaplan, 439–40

- data analysis, 452–53; advice, 694–95; causal flow charts, presentation of results in, 460–62; constant validity check, 453–54; critical thought and statistics, balance of, 647–48; data matrix, 454–56, 497, 498–99, 500; matrices and tables, presentation of results in, 456–60. *See also* qualitative data analysis; quantitative data analysis
- data analysis, challenges, 637–47; measurement and statistical assumptions, 637–38; outlier problem, 569, 571–72, 638–40; shotgun approach, 644–47; statistical significance tests, 640–44
- database management software, 408–12
- data matrix, 454–56, 497, 498–99, 500
- data record, 552
- Davis, Philip, 472
- debriefing, 76, 141–42
- deception, 141; grades of, 441–44
- decision tables, 527–30
- deduction, 9
- deductive coding, 493
- Defense, U.S. Department of, 74, 75
- deference effect, 241–42, 259
- degrees of freedom, 588; *t*-test, two sample, 597
- Democritus, 6
- demographic transition theory, 87–88
- dendrogram, 523, 524, 525, 526
- dependent variables, 34–35
- Descartes, René, 9
- descriptive analysis, 549
- descriptive notes, 397–98
- Dewey, Thomas, 188
- DFA (discriminant function analysis), 692–94
- Dialogue Concerning the Two Chief World Systems* (Galileo), 8
- diaries: field notes and, 391–92; survey research and, 293–95
- dichotomous variables, 30–33
- dietary recall interviews, 297, 418
- diffusion of responsibility, 131
- diffusion of treatments, 117–18
- digital recorders, 228–29
- Dillman's total design method, 280–86
- Dilthey, William, 21
- dimensions, 688
- directional, 584
- direct observation, 413; pros and cons, 435–37; reactivity and, 354, 413, 424–25, 426–27. *See also* continuous monitoring; experience sampling; time allocation studies
- discourse analysis, 473, 484–91; conversation analysis, 486–91
- Discourse on Method* (Descartes), 9
- discovery phase, 493
- discriminant function analysis (DFA), 692–94
- disguised field observation, 437–44; ethics of, 438–44; passive deception, 443–44
- dispersion, measures of, 580–84; interquartile range, 581; range, 581; standard deviation, 581–83; variance, 581–82
- disproportionate stratified random sampling, 155–56
- dissimilarities, 683
- dissimilarity matrix, 456, 683, 690
- distinctive features, 541
- distribution, 169; bimodal distribution, 171, 172, 575, 579, 579–80, 580; leptokurtic distribution, 172, 172, 580; normal distribution, 169, 169, 171, 575; platykurtic distribution, 172, 172, 580; skewed distribution, 172, 172, 575; *t* distribution, 180–81
- distribution, visualization of, 572–80; bar graphs and pie charts, 572, 573; basic shapes, 574–75, 575; bimodal distributions, 579, 579–80, 580; box-and-whisker plots, 576–77, 577; histograms and frequency polygons, 577, 578, 579, 580; stem-and-leaf plots, 573–74, 574

- DLW (doubly labeled water) technique, 297
- Doc (key informant), 197–98
- documentation resources, 96–106; abstracts, 101; meta-analysis, 107–8; online databases, 97–101, 102–5; people, 96–97; review articles, 97; websites, 105–6
- double-barreled questions, 276
- doubly labeled water (DLW) technique, 297
- Dow-Jones Industrial Average, 320
- dowry deaths, 83
- “Driving to Work” (Wallace), 365
- drop-and-collect technique, 258
- dualism, 9
- dummy variables, 496, 623
- echo probe, 218–19
- ecological fallacy, 51–52
- EDMs. *See* ethnographic decision models
- educational model of social change, 35, 86–87
- Educational Resources Information Center (ERIC), 102
- elaboration method, 649–58; further steps, 656–58; tables, building of, 650–54; tables, reading of, 654–56
- ’Elota (Big Man of the Kwaio), 472
- emergent properties problem, 89–90
- empiricism, 3, 4–5
- Enlightenment, 11
- environment, 90
- epistemology, 3
- Equal Employment Opportunity Commission, 438
- equivalence, 547
- Ergas, Savas, 216–17
- ERIC (Educational Resources Information Center), 102
- An Essay Concerning Human Understanding* (Locke), 11
- eta-squared, 621, 635, 637
- ethics, 25–27, 32, 74–78; direct observation, 437; disguised field observation, 438–44; field experiments, 140–43; probing and, 223; rapport and, 369–70; societal codes and, 77; unobtrusive observation, 448
- ethnographic decision models (EDMs), 522–34, 524, 525, 526; building of, 523–25; decision tables, 527–30; IF-THEN charts, 531–34, 533; testing of, 525–27; Young and Garro’s decision tables, 527–30
- ethnographic interviewing, 211. *See also* unstructured interviewing
- Ethnology*, 518
- ethnoscience, 534
- ethogram, 414
- ethology, 414
- event calendars, 248
- events, 463
- evolution problem, 89
- exemplar quote technique, 503–5
- exemplars, 492
- exhaustivity, 46, 511
- expectancy effect, 244–45
- experience sampling, 156, 432–34. *See also* direct observation
- experimental conditions, 111
- experimental method, 109
- experiments, 110–13; classic experiment, 110–12; comparative field experiments, 138–40; confounds, 113, 114–18; field experiments, 110, 131–33, 140–43; laboratory experiments, 110, 130–31; natural experiments, 133–35; naturalistic experiments, 135–40; quasi-experiments, 110; small-world experiments, 136–37; thought experiments, 128–30; true experiments, 110–13. *See also* research designs; threats to validity
- explicit awareness, 364–65
- exploratory phase, 493
- exploratory research, 110
- ex post facto design, 124, 124–25

- external states, 90
external validity, 113
EZ-Text (software), 521
- faces scale, 340, 340–41
face-to-face interviews, 252, 256–58
face validity, 56
factor analysis, 330, 675–81; case example, 678–81; testing for unidimensionality, 335–36
factorial design, 124, 143–45, 144, 167, 236
factorial survey, 291–92
factor loading, 335, 677
farthest-neighbor analysis, 690, 691, 692
Feigl, Herbert, 19
Fei Xiaotong, 376
Fermat, Pierre, 16
FES (focused ethnographic study method), 352–53
field experiments, 110, 131–33; ethics and, 140–43
field experiments, comparative, 138–40
field notes, 387–412, 399, 402, 407; analyzing, 406–12; coding, 398–405, 399, 402, 407; diaries, 391–92; formal field notes, 395–98; jottings, 389–91; logs, 392–94
field notes, analysis of, 406–12; database management, 408–9; relational database systems, 410–12, 411; text management, 406–7
field notes, formal, 395–98
fieldwork. *See* participant observation
Fifi effect, 139
filter questions, 272
finite population correction, 183
first demographic transition, 86
first-order correlation, 660
first quartile, 563
Fisher, Sir Ronald, 601
Fisher's exact test, 614–15
fixed-choice questions, 268–70
flat files, 409, 409
- flow charts: causal flow charts, 460–62; contingency flow charts, 272, 273
focused ethnographic study method (FES), 352–53
The Focus Group (Morgan and Krueger), 237
focus groups, 232–39; case examples, 234–36; data analysis, 238–39; moderating, 237–38; popularity, 233; size, composition, and number, 237; validity, 236–37
folk taxonomies, 522, 534–40, 535; branching tree diagrams, 539, 540; building of, 537–39; covert categories, 534, 535, 536, 539; cultural domains and, 536–37
Ford, Clellan S., 515
forward telescoping, 248
four-fold table, 605, 605
frame elicitation technique, 305–7, 537–38
Framingham Heart Study, 289
F-ratio, 600–601
free listing, 301–5
Freeman, Carla, 190
free pile sorts, 311–15, 315; lumper-splitter problem, 313; networks and, 314–15; objects and, 313–14; P-3 game, 312; taxonomic trees and, 314
French Academy of Science, 10
frequency curves, 173, 577, 578, 579, 580
frequency distribution, 558–60, 558–61; mean calculation, 566
frequency polygons, 173, 577, 578, 579, 580
Freud, Sigmund, 82
Frost, Robert, 19
- Galileo, 8
Galton, Francis, 516
Galton's Problem, 516
gamma, 615–19, 621; test of statistical significance, 618–19

- Garbage Project, 446–48
 Geisteswissenschaften, 21, 81
 Gemeinschaften, 89
 gender: as dichotomous variable, 32–33;
 participant observation and, 373–74
 gender harassment, 384–86
General Enquirer (software), 520
 General Social Survey (GSS), 291–92
 Genovese, Kitty, 131
 Gesellschaften, 89
 Gibson, Gordon, 401, 403
 Gilbreth, F. B., 17
 Gillett-Netting, Rhonda, 290
The Gold Coast and the Slum (Zor-
 baugh), 346
 gold standard test, 58
 Gossett, W. S., 180
 Government Communications Commis-
 sion, U.S., 507
 grades of deception, 441–44
 grand tour question, 220
 Griffith, David, 193
 grounded theory text analysis, 473, 492–
 505; coding, 492–96, 497, 498–99;
 exemplar quote technique, 503–5;
 finding themes, 493–95; memoing,
 492, 497, 500–503. *See also* text anal-
 ysis
 grouped data: mean calculation, 567–68;
 median calculation, 564–66
 group interviews, 232
 GSS (General Social Survey), 291–92
 Gutenberg, Johannes, 7
 Guttman scale, 321–27, 322, 325; case
 examples, 324–26; cautions, 326–27;
 coefficient of reproducibility, 323–24
- Hancock, John, 10
 hanging out, 368–69
 hard-to-find population studies, 190, 191,
 192
 Harris, Marvin, 423–24
 Hayek, F. V., 20
 Helmreich, Robert, 44
 hermeneutics, 22–23, 473–75
 Herskovits, Melville, 199, 426
 heterogeneity of populations, 167
 Hill, Kim, 377
 histograms, 577, 578
 history confound, 114
 Honigman, John, 401
The Hot Pepper Story, 478
 House, Mark, 441
 HRAF. *See* Human Relations Area Files
 humanism, 3, 21–22
 Human Relations Area Files (HRAF),
 463, 515, 516
 Hume, David, 4, 5
 Hunt, George, 465
 Husserl, Edmund, 24
 hypothesis testing, logic of, 584–87; lev-
 els of significance, 585–86; Type I and
 II errors, 586–87
 hypothetico-deductive model of science,
 9
- IBSS (*International Bibliography of the
 Social Sciences*), 102
 ICSSID (International Committee on
 Social Science Information and Docu-
 mentation), 102
 idealist paradigm, 79
 idiographic theory, 81, 83–84
 IF-THEN charts, 531–34, 533
 images, 463
 implications, 547
 independent variables, 34–35
 in-depth interviewing, 210, 212. *See also*
 interviewing
 index cases, 195
 indexes, 320–21. *See also specific*
 indexes; scaling; composite measures
 indicators, 35
 indigenous research, 373
 indirect observation, 413; archival
 research, 448–50; behavior trace stud-
 ies, 444–48. *See also* unobtrusive
 observation

- individual attribute data, 146–47
 individual proximity matrix, 685, 686
 induction, 9
 inductive coding, 493
 inferential analysis, 549
 informal interviewing, 211
 informants, 196–207; cultural competence testing, 203–7, 204–5; finding, 197–99; lies and, 199–200; remuneration of, 209; selection of, 200–203, 206–9
 infrastructure, 87
 instantaneous spot sampling, 425. *See also* time allocation studies
 Institutional Review Boards (IRBs), 143
 instrumental positivism, 20
 instrumentation confound, 115
 intelligence, concept of, 40
 intensive case studies, 190
 interaction chronograph, 415
 interaction effects, 145
 Interaction Process Analysis, 420
 intercoder reliability, 511, 512–15
 interinformant variation, 539
 interitem correlation, 331–32
 internal-external problem, 89
 internal states, 90
 internal validity, 113
International Bibliography of the Social Sciences (IBSS), 102
 International Committee on Social Science Information and Documentation (ICSSID), 102
 Internet-based surveys, 254–55
 interpretive analysis, 22–23, 473–75
 interpretivism, 3, 21–22
 interquartile range, 563, 581
 interrater reliability, 115, 512–15. *See also* intercoder reliability
 interrupted time series design, 128, 128
 intersubjective, 38
 interval variables, 47–48
 intervening variables, 63, 607
 intervention group, 110, 111
 interview control, 210–12
 interviewers, survey research, 265–68; handling of, 268; hiring of, 267–68; training of, 266–67
Interviewer's Manual (University of Michigan), 267
 interviewing, 210–317; informal interviewing, 211; interview control, 210–12; semistructured interviewing, 210, 212. *See also* structured interviewing; unstructured interviewing
 interview schedule, 212
 intracultural variation, 454
 in vivo coding, 402, 404, 493
 IRBs (Institutional Review Boards), 143
 Isaac, Glynn, 376
 ISI (Thompson Institute for Scientific Information), 98
 item analysis, steps to, 330–35; Cronbach's alpha, 332–34; interitem correlation, 331–32; item total correlation, 331, 334–35; scoring the items, 331
 item-total correlation, 331, 334–35
 iterative process, 492

The Jack Roller (Shaw), 346
 Jacobs, Alan, 375
 jargon, participant observation and, 361–62, 364
 Jefferson, Thomas, 12
 John Paul II, Pope, 8
 Johnson, Earvin “Magic”, 128
 Johnson, Jeffrey, 198, 199, 404
 jottings, 389–91
 judgment sampling, 187, 189–91
 Justice, U.S. Department of, 507

 Kadushin, Charles, 228
 Kalymnian case, 66–67
 Kant, Immanuel, 4
 kappa. *See* Cohen's kappa
 Kendall's T_b , 619–20
 Kenny, Michael, 378, 388–89
 key informants, 196–200
 King, Nigel, 222

- kinship systems, 300
 known group comparison, 59
 Kunin, T., 340
 KWIC technique, 495
- Laban, Rudolf von, 420–21
 labanotation, 421
 laboratory experiments, 110, 130–31
 ladder of life, 339, 339–40
La Llorana (Mexican folktale), 483–84
 lambda, 608–9
 lambda 1 design, 310; *See also* BIB
 Lame Deer, John, 472
 landmarks, 249
 language, participant observation and,
 360–63; importance in interviewing,
 224
 least squares line. *See* regression line
 Left Handed (Navajo man), 472
 Leibnitz, Gottfried Wilhelm von, 4
 leptokurtic distribution, 172, 172, 580
 level of analysis, 80
 level of significance, 585–86
 levels of contrast, 538–39
LEXIS/NEXIS, 104
 Likert scale, 327–35; building of,
 328–35; item analysis, 330–35
 Likert-type scales, 328
 linear relations, 603
 linear transformations, 591
 Linguistics and Language Behavior
 Abstracts (LLBA), 104
Literary Digest, 278–79
 literature search. *See* documentation
 resources
 lived experience, 213
 LLBA (Linguistics and Language Behav-
 ior Abstracts), 104
 loaded questions, 276
 location cues, 248
 Locke, John, 4, 11
 logical empiricism, 19
 logical operators, 547
 logical positivism, 19
 logical time order, 61, 63
 logic of experiments, 109
 logic of hypothesis testing, 584–87
The Logic of Modern Physics (Bridg-
 man), 44
 logs, 392–94
 longitudinal design, 288
 longitudinal survey data, 463–64. *See*
 also specific longitudinal surveys
 long question probe, 219–20
 loose talk phenomenon, 304
 lost-letter technique, 137
 Lowe, John, 342
 Lucretius Carus, Titus, 7
 lumper-splitter problem, 313
 Luther, Martin, 7
 Lysenko, T. D., 6
- Mach, Ernst, 18
 magnitude, 584
 main effects, 145
 mall-intercept survey, 161, 257
 MANOVA (multiple-way ANOVA), 602
 marginals, 605
 materialist paradigm, 79
 matrices, 454–60
 maturation confound, 114–15
 maximizing the between-group variance,
 154, 162–66
 MCAPI (mobile computer-assisted per-
 sonal interviewing), 254
 McCarty, Christopher, 253
 MCI, 233
 McNabb, Steven, 377
 MDS. *See* multidimensional scaling anal-
 ysis
 Mead, Margaret, 464
 mean, 562, 566–72; grouped data, calcu-
 lation for, 567–68; large samples, test-
 ing of, 588–89; mathematical feature
 of, 568–69; outlier problem, 569,
 571–72; under various distributions,
 575, 575. *See* central limit theorem
 measurement, 28, 35–37; levels of,
 45–49; rule about, 49

- median, 562, 563–66; grouped data, calculation for, 564–66; under various distributions, 575, 575
- Meditations* (Descartes), 9
- MEDLINE, 103
- memoing, 492, 497, 500–503
- memory building, 365–66
- The Mending Wall* (Frost), 19
- mental maps, 684, 687, 687
- Merton, Robert, 5
- meta-analysis, 107–8, 130
- metaphors, schemas and, 482
- metaphysics, 5–6
- method, meanings of, 3
- methodological notes, 395–97
- metric data, 681
- Mexican folk tale schema, 483–84
- Micturition Study, 442–43
- The Middle Five* (La Flesche), 471
- Milgram's Obedience Experiment, 75–77
- Mill, John Stuart, 11, 13–14, 18
- “A Million Random Digits with 100,000 Normal Deviates” (Rand Corporation), 150
- minidisk recorders, 229–30
- minimizing the within-group variance, 154
- mixed methods, 298, 384–386
- mode, 562–63; under various distributions, 575, 575
- modulus, 181
- Morris, Beverly, 374
- mortality confound, 117
- Mountain Wolf Woman (Winnebago woman), 472–73
- multicollinearity, 661, 672–75
- multidimensional scaling analysis (MDS), 681–89, 684, 687, 688; cognitive world examples, 684–89; physical world example, 683–84
- multidimensional variables, 29–30
- multiple free pile sorts, 312
- multiple indicator scales, 319–20
- multiple-R, 661
- multiple regression, 661–69; application of, 663–67; case examples, 667–69; equation, 661–63; stepwise multiple regression, 667. *See also* path analysis
- multiple-R squared, 661
- multiple-way ANOVA (MANOVA), 602
- multivariate analysis, 549, 550, 649–94; cluster analysis, 689–92; discriminant function analysis (DFA), 692–94; factor analysis, 675–81; multicollinearity, 661, 672–75; partial correlation, 658–61; path analysis, 670–72. *See also* elaboration method; multidimensional scaling; multiple regression
- multivariate relations, 95
- Mumikhpykak, Boris, 377
- Murdock, George Peter, 515
- mutual exclusivity, 46, 511
- Ñähñu native ethnography, 467–70
- naïveté, participant observation and, 366–68
- narrative and performance analysis, 473, 475–79
- National Association of Practicing Anthropologists, 75
- National Black Election Study, 262
- National Technical Information Service (NTIS), 102–3
- native ethnography, 465–73; genres, 471–73; Wolcott's critique of, 468–70
- native ethnography, Ñähñu, 467–70
- natural experiments, 133–35
- naturalistic experiments, 135–40
- natural logarithm, 634
- nature-nurture problem, 89
- Naturwissenschaften, 21, 81
- necessary condition, 548
- negative cases, 500
- negative covariants, 603
- network cues, 248–49
- network sampling, 192–94. *See* sampling, chain referral
- neutral probe, 219

- new error, 608
 Newton, Isaac, 9, 11
 Nisa (!Kung woman), 472
 nominal variables, 45–47
 nomothetic theory, 81, 84–86
 nonaffiliated generics, 536
 nondirectional, 584
 nonlinear relations, 604, 632–35
 nonprobability sampling, 147, 186–96;
 case control sampling, 194–95; convenience sampling, 191–92; credibility and, 195–96; network sampling, 192–94; purposive sampling, 187, 189–91; quota sampling, 187–89; uses, 186–87. *See also* informants; sampling
 nonreactive observation, 413
 nonspecific prompting, 302
 normal distribution, 169, 169, 171, 575
 norm referenced, 320
 Nosnibor effect, 51–52
Notes and Queries on Anthropology (Royal Anthropological Institute), 345–46
 NTIS (National Technical Information Service), 102–3
 null hypothesis, 584
 Nutrition Knowledge and Attitude Questionnaire, 120
 NVivo (software), 521
- objectivity, participant observation and, 370–73
 observables, measurement of, 48
 observing participant, 347, 348
 obtrusive observation. *See* direct observation
 Ockham's razor, 59
 OCLC (Online Computer Library Center), 105
 OCM (*Outline of Cultural Materials*), 400, 400–401, 402, 468, 515
 odds ratio, 613
 Ogburn, William F., 20
- old error, 608
 Olopinikwa (Kuna chief), 477
 one-group pretest-posttest design, 125–26, 126
 one-shot case study design, 124, 124–25
 one-tailed test, 584
 one-way ANOVA, 602
 Ongka (Big Man of the Kawelka), 472
 Online Computer Library Center (OCLC), 105
 online databases, 97–101, 102–5; search tips, 106–7
On the Nature of the Universe (Lucretius), 7
On The Social Contract (Rousseau), 12
 open coding, 493
 open-ended questions, 268–70
 operational definitions, 38, 41–45; pros and cons, 42–45
 operationalize, 28
 operational notes, 499, 500
 order effects, 315
 ordinal variables, 47
 ordinate, 572, 603
 orthogonal solution, 677
 Oths, Kathy, 135
 outlier problem, 569, 571–72, 638–40
Outline of Cultural Materials (OCM), 400, 400–401, 402, 468, 515
- Paine, Thomas, 12
 paired adjectives, 336
 paired comparisons, 315–16, 316
 paired objectives, 336
 panel design, 288
 panel studies, 149, 288–90; attrition and, 291
 Panel Study on Income Dynamics (PSID), 288
 paradigms, 79–81; consequences of, 86–88
 parameter estimators, 167
 parent population, 595–96
 Park, Robert, 346

- parsimony, principle of, 59
- partial correlation, 658–61
- partial out, 660
- participant observation, 342–86; defined, 343–44; background and history, 344–47; field entry, 356–59; fieldwork roles, 347–49; gender and, 373–74; parenting and, 375; personal characteristics and, 375; sex and, 375–76; survey research, mixed with, 384–86; survival, 376–78; time investment, 349–53; validity and, 354–56. *See also* unobtrusive observation
- participant observation, stages of, 378–84; break, 382; culture shock, 379–81; departure, 383–84; exhaustion, 383; focusing, 382–83; initial contact, 378–79; obvious, discovery of the, 381–82
- participant observation skills, 359–73; explicit awareness, 364–65; jargon, 361–62, 364; language, 360–63; memory building, 365–66; naiveté, 366–68; objectivity, 370–73; rapport, 368–70; writing, 368
- participant observer, 347
- participatory mapping, 352
- participatory rapid assessment (PRA), 352
- participatory transects, 352
- Pascal, Blaise, 16
- passive deception, 443–44
- path analysis, 670–72. *See also* multiple regression
- path coefficients, 671
- Pearson's product moment correlation (r), 621, 622–24, 629–30; shortcut formula, 626; testing the significance of, 630–32
- Pelley, William Dudley, 507
- periodicity, 151–52
- phased assertion, 221–22
- phenomenology, 23–24
- phonology, 480, 541
- Piaget, Jean, 82
- pie charts, 572, 573
- Pika, Alexander, 377
- pile-sorting, 494–95. *See also* free pile sorts
- pilot studies, 190
- Plato, 4
- platykurtic distribution, 172, 172, 580
- pooled variance, 596–97
- Popol Vuh* (Quiché Maya manuscript), 476–77
- population, 173
- population parameters, 146
- positive covariants, 603
- positivism, 3; alternatives to, 20–24; early, 12–15; instrumental positivism, 20; logical positivism, 19; social activism and, 18; Vienna Circle, 18–19
- positivist school, 13
- postdicting, 66, 534
- posttest-only with random assignment design, 123, 123–24
- posttests, 120
- power, statistical testing and, 642, 644
- power analysis, 587
- PPS sampling. *See* probability proportionate to size sampling
- practical significance, 642
- PRA (participatory rapid assessment), 352
- Prebisch, Raul, 39
- precision, 54
- predictive validity, 59
- PRE (proportionate reduction of error), 607–9
- pretests, 111–12, 120; questionnaires and, 286–87
- primitive axioms, 65
- Principles of Professional Responsibility* (AAA), 74
- printing press, 7
- probability proportionate to size (PPS) sampling, 159–66; field techniques, 160, 160–62; maximizing between-group variance, 162–66

- probability sampling, 147
 probability theory, 15–16
 probability value, 585, 640
 probing, 217–23; baiting, 221–22; echo probe, 218–19; ethics of, 223; long question probe, 219–20; neutral probe, 219; probing by leading, 220–21; silent probe, 218; tell-me-more probe, 219; types of, 302
 profile analysis, 454
 profile matrix, 454–55, 455
 Project Camelot, 74
 proportionate reduction of error (PRE), 608–9
 proportionate stratified random sample, 154
 proportion of matching pairs, 616
 proportions, estimation of, 181–84
 Protagoras, 21
 Protestant Reformation, 7
 proxemics, 441
 proximity matrix, 454, 455, 456, 456, 684
 pseudopatients and pseudoclients 437
 PSID (Panel Study on Income Dynamics), 288
 PsycARTICLES, 103
 PsycINFO, 103–4
 published codes, 518
 purposive sampling, 187, 189–91. *See* sampling, judgment sampling
p-value, 585, 640
Pygmalion in the Classroom (Rosenthal and Jacobson), 244

 QCA (qualitative comparative analysis), 545
 QDA Miner (software), 521
 QFF (Quantitative Food Frequency) questionnaire, 297
 qualitative comparative analysis (QCA), 545
 qualitative data analysis, 451–52; analytic induction, 522, 544, 548; componential analysis, 522, 541–44, 542, 543. *See also* data analysis; ethnographic decision models; folk taxonomies; text analysis
 qualitative measurement, 46
 qualitative variables, 556
 quantitative data analysis, 451–52. *See also* bivariate analysis; data analysis; multivariate analysis; univariate analysis
 Quantitative Food Frequency (QFF) questionnaire, 297
 quantitative measurement, 47
 quasi-experiments, 110
 questionnaires, 251–98; computers and, 252–55; cross-sectional surveys, 288; data collection methods, 252, 298; data collection methods, choosing, 264–65; dietary recall interviews, 297, 418; face-to-face interviews, 252, 256–58; factorial surveys, 291–92; Internet-based surveys, 254; interviewers and, 265–68; longitudinal surveys, 288–89; nonresponse, adjusting for, 279–80; panel studies, 288–91; participant observation, mixed with, 384–86; question wording and format, 270–77; randomized response, 295–97; remuneration and, 284–85; response rates, 278–86; self-administered questionnaires, 252, 258–61; telephone interviews, 261–64; time budget surveys and diaries, 293–95
 questionnaires, building of, 268–87; back translation, 277–78; closed vs. open-ended questions, 268–70; pretesting, 286–87; response rates, improving, 280–86; rules, 270–77
 Quételet, Adolphe, 11, 12, 18
 quota sampling, 187–89

 race, as dichotomous variable, 30–32
 Radin, Paul, 466
 Rand Corporation, 150
 random assignment, 111

- random digit dialing (RDD), 261–62, 263
randomized response technique, 295–97
range, 581
rank ordering, 316–17
rapid assessment procedures, 352–53
rapport, participant observation and, 368–70
rating scales, 317
rationalism, 3–4
ratios, 563
ratio variables, 48
RDD (random digit dialing), 261–62, 263
reason, 3, 9
reactivity: direct observation and, 354, 413, 424–25, 426–27; structured interviewing and, 257
recalibration, 55
recording equipment, 228–32
Redfield, Robert, 346
redundant question probe, 302
regression, 625–29
regression line, 624, 625, 626–27, 636
regression to the mean, 115–16
relational database systems, 410–12, 411
relative deprivation theory, 65, 66
reliability, 6, 28, 54
remuneration: of informants, 209; questionnaires and, 284–85
research designs, 109, 118–28, 119; factorial design, 124, 143–45, 144, 236; interrupted time series design, 128, 128; one-group pretest-posttest design, 125–26, 126; one-shot case study design, 124, 124–25; posttest-only design with random assignment, 123, 123–24; Solomon four-group design, 121, 121–22; two-group posttest-only design without random assignment, 126, 126–28; two-group pretest-posttest design without random assignment, 122, 122–23; two-group pretest-posttest design with random assignment, 118–21, 120. *See also* experiments
researcher response, 378
research hypothesis, 584
research process, 69–74; ideal, 69–70; money, 73–74; people, 74; personal interest, 70–71; resources, 72–74; science vs. nonscience, 71–72; time, 72–73
research topics, types of studies, 90–95
residuals, 628, 663
respondent-driven sampling (RDS), 192, 193–94
respondent mortality, 291
response effects, 89, 239–45, 259; deference effect, 241–42, 259; expectancy effect, 244–45; third-party-present effect, 242–43; threatening questions, 243–44
restricted time, 249–50
review articles, 97
Rice, Stuart, 239
Richards, William, 377
Rivers, W. H. R., 109
robust, 321
role cues, 249
Rosaldo, Michelle Zimbalist, 376
Rosenbaum, Harvey, 466
Rousseau, Jean-Jacques, 11–12
routine activity theory, 64
Royal Anthropological Institute, 100
R-squared, 661
r-squared, 623, 629–30

Saint-Simon, Claude-Henri de, 11, 12–13
Salinas Pedraza, Jesús, 197, 467
Samoan Energy Study, 418
sample accuracy, 166
sample precision, 166, 168
sample size: determinants, 166–68; means, calculation for estimating, 179; proportions, calculation for estimating, 181–84; smaller populations, calculation for estimating proportions, 183–84
sample statistics, 167

- sampling, 146–68; case control, 194–95; cluster sampling, 157–58; chain referral, 192–94; complex sampling designs, 158; convenience sampling, 191–92; cultural data, 147, 187; judgment sampling 189–91; obstacles to, 148–49, 151–52; periodicity, 151; probability proportionate to size (PPS) sampling, 159–66; probability vs. non-probability sampling, 147; quota sampling, 187–89; simple random sampling, 149–51; space sampling, 160–62; stratified random sampling, 153–56; systematic random sampling, 151–53; time allocation studies and, 427–31. *See also* nonprobability sampling
- sampling by convenient replacement technique, 257
- sampling design, 154
- sampling distribution, 173
- sampling frames, 149
- sampling from a telephone book, 152
- sampling interval, 151
- sampling theory, 169–85; calculation for sample size for estimating proportions, 181–84; central limit theorem, 173–78; confidence intervals, 178–79; distribution, 169–73; *t*-distribution, 179–81
- Sapir, Edward, 465–66
- Sarkar, N. K., 355
- scaling: faces scale, 340, 340–41; Guttman scale, 321–27; ladder of life, 339, 339–40; Likert scales, 327–35; multiple indicator scales, 319–20; semantic differential scaling, 274, 336–38, 337; single indicator scales, 318–19, 320. *See also* indexes; composite measures
- scattergrams, 603, 604, 635, 636, 639
- scatterplots, 603, 604, 635, 636, 639
- schema analysis, 473, 479–84; case examples, 481–84
- schemas, 480–81
- Schiller, Ferdinand C. S., 21
- Schutz, Alfred, 24
- SCI (*Science Citation Index Expanded*), 98, 99
- science: development of, 6–10; norms, 5–6
- Science Citation Index Expanded* (SCI), 98, 99
- scientific management, 17
- scientific method, 6, 17; assumptions, 5
- scientism, 20
- scratch notes, 389–91
- scripts, 480–81
- SE (standard error), 587
- second demographic transition, 86
- second-order correlation, 660
- segmenting, 510
- selection bias, 116–17
- self-administered questionnaires, 252; pros and cons, 258–61
- self-anchoring scale, 339
- Seligman, C. G., 375
- SEM (standard error of the mean), 587
- semantic cues, 302
- semantic differential scaling, 274, 336–38, 337
- semantic network analysis, 495
- semantics, 480
- semistructured interviewing, 210, 212. *See also* interviewing
- sentence frame technique, 305–7, 537–38
- sex, fieldwork and, 375–76
- sexual harassment, 384
- shaman effect, 209
- shape of distributions, 572–80
- Shopping Study, 418–19
- shotgun approach, 644–47
- silent probe, 218
- SIL (Summer Institute of Linguistics), 396
- similarities, 681
- similarity matrix, 206, 456, 685, 686
- simple matching coefficient, 679
- simple random sampling, 149–51

- simple scales, 318–19, 320
 “Since the Eruption of Mt. St. Helens”
 (Loftus and Marburger), 249
 single indicator scales, 318–19, 320
 single-link clustering, 690–91, 692
 Six Culture Study, 422–23
 skepticism, 4
 skewed distribution, 172, 172, 575
 Skinner, B. F., 82
 small-world experiment, 136–37
 snowball sampling, 187, 192, 193. *See*
 sampling, chain referral
 social activism, positivism and, 18
 social contact theory, 64
 social contract, 11–12
 social desirability effect, 250
 social facts problem, 89–90
 social indicators, 569
Social Physics (Quételet), 18
 social research: cause and effect, 28,
 60–66; as a craft, 1–2; measurement,
 28, 35–37, 45–49; qualitative vs. quan-
 titative, 24–25; 451–52; reliability, 28,
 54; sameness within social sciences,
 2–3. *See also* research designs; theory;
 validity; variables
 social sciences: development of, 10–12;
 success and failures of, 15–18
Social Sciences Citation Index (SSCI),
 98, 99
 Society for Applied Anthropology, 75
 sociobiological paradigm, 79
 socioeconomic status (SES), 37, 319
Sociological Abstracts, 103–4
 sociology, naming controversy, 14
 Solomon four-group design, 121, 121–22
 Spearman, Charles E., 676
 Spearman’s r , 621–22
 specialized informants, 196, 200–202,
 206–7
 Spence, Janet, 44
 Spier, Leslie, 343
 split-half reliability test, 333
 spot sampling, 413, 425. *See also* time
 allocation studies
 spuriousness, 61, 62–63
 Stalin, Josef, 6
 Standard Cross-Cultural Sample of 186
 societies (Murdock and White), 516
 standard deviation, 170, 581–83
 standard error (SE), 170, 587
 standard error of the mean (SEM), 587
 standardized coefficients, 666
 standardized values, 671
 standard scores, 171, 590–91
 static group comparison, 126–28
 statistical power, 642–44
 statistical significance, 585–86, 595, 642
 statistical significance testing, 595; chi-
 square, 592; F -ratio 601; gamma,
 618–19; Pearson’s r , 630–32; t -test,
 two sample, 597–98; validity of,
 640–44
 stem-and-leaf plots, 573–74, 574
 stepwise multiple regression, 667
 stimulus group, 110, 111
 strategic methods, 3, 343–44
 stratified random sampling, 153–56
Street Corner Society (Whyte), 197
 street-intercept survey, 161, 257
 string variables, 555
 structure, 87
 structured interviewing, 212, 251–317.
 See also cultural domain analysis;
 interviewing; questionnaires
The Subjection of Women (Mill), 18
 subjectivity, 22
 sucker bias, 361
 sufficient condition, 548
 Summer Institute of Linguistics (SIL),
 396
 sum of the squares, 598, 599–600
 superstructure, 87
 Supreme Court, U.S., 147, 438
 survey research. *See* questionnaires
SYMLOG (Bales and Cohen), 420
 symmetric, 685
 synoptic problem, 474
 syntax, 480

- systematic bias, 55, 111
 systematic random sampling, 151–53
System of Positive Policy (Comte), 13
- tabula rasa, 4
 Talayesva, Don, 472
 talk-in-interaction, 487
 Tambiah, S. J., 355
 target item, 336
Tau_b. See Kendall's *T_b*
 taxonomic analysis. See folk taxonomies
 taxonomic trees, 314
 Taylor, Frederick Winslow, 17
t distribution, 180–81; critical value, 181
 Tearoom Trade Study, 439
 telephone interviews, 252; pros and cons, 261–64
 tell-me-more probe, 219
 testing confound, 115
 text analysis, 463–64, 473–521; computers and, 519–21; discourse analysis, 484–91; hermeneutics, 473–75; narrative and performance analysis, 475–79; schema analysis, 479–84. See also content analysis; grounded theory text analysis
 text collection, 464–65. See also native ethnography
 text management software, 406–7
 Tezcatlipoca (Mesoamerican god), 508
 TFRs (total fertility rates), 86, 571
 themes, 492–95
 theoretical perspectives. See paradigms
 theoretical saturation, 500–501
 theory, 28, 61, 63–66, 78–89; idiographic theory, 83–84; idiographic vs. nomothetic, 81–83; nomothetic theory, 84–86; paradigms, 79–81, 86–88
 theory notes, 499, 500
 thinkaloud method, 286–87
 third-party-present effect, 242–43
 third quartile, 563
 Thompson Institute for Scientific Information (ISI), 98
- thought experiments, 128–30
 threatening questions, 243–44
 threats to validity, 109; confounds, 114–18
 threats to validity, controlling for, 118–28; interrupted time series design, 128, 128; one-group pretest-posttest design, 125–26, 126; one-shot case study design, 124, 124–25; posttest-only design with random assignment, 123, 123–24; Solomon four-group design, 121, 121–22; two-group posttest-only design without random assignment, 126, 126–28; two-group pretest-posttest design without random assignment, 122, 122–23; two-group pretest-posttest design with random assignment, 118–21, 120
 time allocation studies, 425–32; coding, 431–32; continuous monitoring, mixing with, 434–35; reactivity, 426–27; sampling, 427–31. See also direct observation
 time budget surveys, 293–95
 timeline cues, 249
 time sampling, 425. See also time allocation studies
 topical codes, 399–404
 Tosmana (software), 547
 total fertility rates (TFRs), 86, 571
 total variance, 598, 599, 600
 Tozzer Library (Harvard University), 100, 101
 trained subjectivity, 22
 transcribers, 230–31
A Treatise on Man (Quételet), 12
 treatment group, 110, 111
 triad tests, 307–11; BIB design, 309–11, 310
 Trobriand land dispute schema, 481–82
 true experiments, 110–13. See also field experiments; laboratory experiments
 Truman, Harry S, 188
t-distribution, 179–81

- T*-scores, 591
t-test, two sample, 595–98; test for statistical significance, 597–98
 24-hour recall test, 297, 418
 two-group posttest-only design without random assignment, 126, 126–28
 two-group pretest-posttest design without random assignment, 122, 122–23
 two-group pretest-posttest with random assignment, 118–21, 120
 Two Leggings (Crow man), 472
 two-tailed test, 584
 Tylor, Edward, 516
 Type I error, 586–87
 Type II error, 586–87
- unconstrained pile sort method. *See* free pile sorts
 unidimensional variables, 29–30, 321, 335
 unique identifier, 552
 unitizing, 510
 units of analysis, 49–52; ecological fallacy, 51–52; rule about, 51
 univariate analysis, 549; chi-square test, 591–92, 593; coding and, 550–55; dispersion measures, 580–84; distribution, visualization of, 572–80; frequency distribution, 558–60, 558–61, 566; raw data, 555–58
Unobtrusive Measures (Webb), 445
 unobtrusive observation: disguised field observation, 437–44; ethics, 448. *See also* participant observation; indirect observation
 unstandardized regression coefficients, 666
 unstructured interviewing, 210, 211–12, 213–17; case example, 213–15; initial steps, 215; language, importance of, 224–25; pacing, 225; practice, 223–24; presentation of self, 226–27; probing, 217–23; respondents, handling of, 216–17, 222–23; voice recording and, 227–32. *See also* interviewing
Urban Life and Culture, 346
- validity, 28, 53; determination of, 56–60; of focus groups, 236–37; internal vs. external validity, 113; participant observation and, 354–56; problem with, 60; of statistical significance testing, 640–44. *See also* threats to validity
 value neutrality, 372
 values, 35–37
 variables, 28–29; antecedent variables, 63, 607; character variables, 555; conceptual definitions of, 37–40; dependent vs. independent, 34–35; dichotomous variables, 30–33; dummy variables, 496, 623; interval and ratio variables, 47–48; intervening variables, 63, 607; major kinds of, 90–91; multidimensional variables, 29–30; nominal variables, 45–47; ordinal variables, 47; qualitative variables, 556; transforming variables, 634–35; unidimensional variables, 29–30, 321
 variance, 581–82; accounting for, 630–31
 video, 423–25
 Vienna Circle of logical positivists, 18–19
 vignettes, 291–92
 voice recognition software, 231–32
 Voltaire, 11
- Wah'Kon-Tah* (Mathews), 471
 Watergate experiment, 142–43
The Way to Rainy Mountain (Momaday), 471–72
WCJ (*World Cultures Journal*), 518
Web of Science, 97, 98–100
 weighting, 156, 663
 Whyte, W. F., 197

- Wentz, Frances Thundercloud, 473
Werner, Oswald, 299
Wichita Study, 163–65
Wilder, Douglas, 241
William of Ockham, 59
Williams, Holly, 230
Williams, Reverend Samuel, 10
Windelband, Wilhelm, 81
Wisconsin Longitudinal Study, 289
withholding of treatment, 141
within-group variance, 598, 599, 600;
 minimizing, 154
word lists, 495
World Cultures Journal (WCJ), 518
world systems theory, 65
Wright, Sewall, 670
writing, participant observation and, 368
 χ^2 . *See* chi-square
Xiloj, Andrés, 476–77
Ximénez, Francisco, 476
Young and Garro's decision tables,
 527–30
Yule's *Q*, 620–21
Zapotec children Study, 415–18
zero-order correlation, 659, 660
zero point, 48
z-scores, 171, 590–91
Zwickler, Thomas, 377

Author Index

Numbers in italics refer to figures and tables.

- Abel, M. H., 92
Abelson, R. P., 480
Adair, J., 467
Adams, K., 85
Addams, J., 18
Agar, M., 198, 359, 367
Ahrentzen, S., 61
Aiello, J. R., 441
Akkerboom, H., 283
Aldridge, D., 2
Alexander, J., 300
Algase, D. L., 414
Al-Krenawi, A., 192
Allen, J. T., 495
Allport, F. H., 449
Al-Nuaim, A. A., 158
Altmann, J., 417
Altorki, S., 374
American Statistical Association, 281
Anderson, A. B., and E. S. Anderson, 158
Anderson, C. A., 64
Anderson, M., 465, 466
Andrews, F. M., 340
Angrosino, M. V., 343
Aquilino, W. S., 243, 259
Archer, D., 445
Aronson, E., 130
Arp, W., III, 93
Ashkenazi, M., 376
Ashton-Vouyoucalos, S., 127, 504–5,
506
Associated Press, 424
Aunger, R., 50–51, 240
Axinn, W. G., 240, 267
Ayres, I., 438
Bachman, J. G., 283
Baker, J., 94
Baker, R., 211, 232
Baksh, M., 410–11
Bales, R. F., 420, 421
Balicki, A., 466–67
Barber, N., 518
Barker, J. C., 343
Barnes, J. H., 253
Barroso, J., 190
Barry, H., III, 61, 85–86, 518
Bartlett, F. C., 109
Bartlett, K., 444
Basiotis, P. P., 91
Bassuk, E. L., 194–95
Baumgartner, R., 285
Beautrais, A. L., 194
Bechky, B. A., 93
Bechtel, R. B., 414
Bechtel, R. J., 515
Beck, A. T., 259
Beck, C. T., 278
Becker, H. S., 346
Begley, S., 264
Belk, R. W., 94, 195
Belo, J., 465
Belter, R. W., 341

- Bem, S. L., 32
 Bercovici, S., 290
 Berk, R., 291
 Berman, J., 465
 Bermant, G., 10
 Bernard, H. R., 3, 55, 64, 67, 71, 91, 127,
 157, 161, 214, 291, 347, 368, 389, 401,
 429, 467–68, 469–70, 478–79, 479,
 494, 506, 552, 556, 589, 596, 616, 693
 Berra, Y., 413
 Berreman, G. D., 350
 Berry, J., 327
 Bessinger, J. B., 495
 BeVier, C. S., 281
 Bickel, R., 35
 Bielawski, D., 92
 Bieri, D., 341
 Billiet, J., 266
 Birdwell-Pheasant, D., 458, 459
 Black, D., 414
 Blair, E., 269
 Blalock, H. M., Jr., 640
 Boas, Franz, 360
 Bochner, S., 139–40
 Boeckelman, K., 93
 Boelen, W. A. M., 198
 Bogdan, R., 366, 387
 Bohannan, Paul, 469
 Bolton, R., 387
 Booth, C., 18
 Borgatti, S. P., 299, 301, 310, 314, 326,
 547, 556, 589, 616, 683
 Borgerhoff-Mulder, M. B., 427
 Boruch, R. F., 122
 Boserup, E., 84–85, 108
 Boster, J. S., 56, 85, 202, 305, 311, 313,
 314
 Bourgois, P. I., 343, 347–48, 354, 372
 Bowman, K. H., 153
 Boyle, E., Jr., 638
 Bradburn, N. M., 239, 247, 248, 249, 260,
 269, 272, 282, 296
 Bradley, C., 377–78
 Brent, E. E., 414
 Brettell, C. B., 448
 Brewer, D. D., 248, 249, 301–2
 Bridge, R. G., 123–24
 Bridgman, P. W., 44
 Brief, A. P., 341
 Briere, J., 93
 Brislin, R. W., 277, 361
 Britt, C., 128
 Brock, J. K.-U., 258
 Browne, K. E., 190
 Brugge, D. M., 148
 Bryant, C., 20
 Burling, R., 360, 544
 Burton, M. L., 309, 310, 310, 311
 Butler-Smith, P., 512

 Cahn, P. S., 290
 Cahnman, W. J., 152
 Cameron, S., 512
 Campbell, D. T., 114, 122, 123, 134–35,
 277
 Campbell, G. R., 449
 Cannell, C. F., 217, 239, 247, 270
 Cantril, H., 339, 339
 Caplan, P. J., 75, 89
 Carey, J. W., 267, 514
 Carlson, R. A., 512
 Carneiro, R. L., 324
 Caro, T. M., 427
 Carver, R. P., 640
 Caserta, M. S., 264–65
 Cassell, J., 375
 Cassidy, R. C., 46
 Casson, R., 480
 Casterline, J. B., 243
 Chadsey-Rusch, J., 414
 Chagnon, N., 379
 Chambers, R., 352
 Chan, A., 545
 Chapkis, W., 440
 Chapman, P., 120
 Chapple, E. D., 415
 Chick, G., 433, 434
 Chidambaram, V. C., 243

- Choi, N., 33
 Chomsky, N., 4
 Cialdini, R. B., 135
 Claiborne, W., 83
 Cliggett, L., 290
 Clignet, R., 65
 Coombs, R. B., 92
 Cobb, W. J., 217, 240
 Cochran, W. G., 183
 Cohen, A., 92
 Cohen, J., 513, 640, 644
 Cohen, S. P., 420
 Cohen-Kettenis, P. T., 46
 Cohn, E. G., 64–65, 634, 634
 Colby, B. N., 520
 Cole, D., 392
 Cole, M., 252
 Coleman, J. S., 17
 Collins, A., 512
 Colson, E., 290
 Comitas, L., 127, 157, 693
 Comte, A., 13
 Conaway, M. E., 374
 Connolly, M., 211
 Conrad, F. G., 256
 Converse, J. M., 219, 222
 Cook, T. D., 123
 Cooke, C. A., 91
 Corbin, J., 402–4, 492, 493, 499
 Corral-Verdugo, V., 93
 Cowan, G., 509
 Crabb, P. B., 92
 Csikszentmihalyi, M., 156, 432
 Curtice, J., 188

 Dadds, M. R., 414
 Daly, M., 79
 D'Andrade, R. G., 247, 304, 308, 364,
 494, 535, 538
 Darley, J. M., 131
 Darwin, C., 414
 Davis, D., 375
 Davis, D. W., 262
 Davis, J. A., 620

 Davis, K., 129–30
 Davis, N. Z., 7
 de Ghatt, V. J., 690, 690
 Dehavenon, A. L., 424
 Dei, Kojo A., 343
 de Leeuw, E. J., 259
 Dellino, D., 62
 Deloria, V., 360
 de Rada, V. D., 281
 Descartes, R., 9
 DeVellis, R. F., 330
 DeWalt, K. M., 374
 Dickerson, S. S., 190
 Dillman, D. A., 261, 280, 283, 284
 Dilthey, W., 21
 Ditmar, H., 94
 Divale, W. T., 519
 Dobkin de Rios, M., 349, 612, 613
 Dohrenwend, B. S., 210, 217
 Donald, G., Jr., 415
 Donnerwerth, G. V., 91
 Doob, A. N., 136
 Dordick, G. A., 72–73
 Doughty, P., 50
 Dow, M. M., 516
 Dressler, W. W., 31, 65–66, 308, 641
 Drickamer, L. C., 414
 Drury, C. C., 414
 Dufour, D., 590–91
 Dugdale, V., 253
 Dugger, C. W., 83
 Dungy, C. I., 317
 Durkheim, E., 89, 638, 673
 Dyk, W., 472
 Dyk, W., and R. Dyk, 472
 Dyl, J., 93

 Easlea, B., 10
 Eckman, P., 414
 Edgerton, R. B., 89, 92, 289–90
 Edmondson, B., 446
 Edmonson, M. S., 478
 Edwards, M., 287
 Eibl-Eibesfeldt, I., 423

- Eisenstein, E., 7
 Ekman, P., 341
 El Guindi, F., 466–67
 Elifson, K., 361–62
 El-Solh, C. Fawzi, 374
 Ember, C. R., and M. Ember, 516, 518
 Emerson, R. M., 391
 Erasmus, C. J., 426
 Erdoes, R., 472
 Erikson, K. T., 440–41
 Erlich, L., 445
 Evans, M. J., 389
 Evans-Pritchard, E. E., 369–70, 375

 Farnell, B., 421–22, 484
 Farrington, J. M., and M. G. Farrington, 495
 Feigl, H., 19
 Feldman, R. E., 138–39
 Fenton, J., 314
 Festinger, L. A., 64
 Fetterman, D., 366
 Fink, J. C., 252
 Finkel, S. E., 241
 Finkler, K., 134
 Fisher, R. A., 692
 Fjellman, S. M., 458, 460, 460
 Fleisher, M., 343, 348
 Flores, E., 305, 444
 Fluehr-Lobban, C., 74, 75
 Foster, G., 290, 401
 Fowler, F. J., 247, 279
 Fox, R. J., 281, 283
 Frake, C. O., 538, 541
 Frank, S. H., 414
 Freeman, D., 360
 Freeman, L. C., 247
 Freilich, M., 198, 389–90
 Frisancho, A. R., 591
 Fry, D. P., 415–18
 Fuqua, D. R., 33

 Gable, R. K., 278
 Gaito, J., 638
 Galilei, Galileo, 8

 Gans, L. P., 694
 Garagiola, J., 413
 Garrett, S. B., 248, 249
 Garro, L. C., 305, 522, 527–30, 528, 529, 530
 Gatewood, J. B., 303–4, 522
 Gaulin, S. C., 85
 Gearing, J., 349
 Geertz, C., 23, 425, 436
 Gerard, H. B., 131
 Gilens, M., 91
 Gilly, M. C., 509–10
 Gittelsohn, J., 425
 Gladwin, C. H., 522, 523
 Gladwin, H., 458, 460, 460, 522
 Glaser, B. G., 492, 493, 503
 Glaser, J. M., 91
 Glass, G. V., 107
 Glazer, M., 439–40
 Godoy, R., 92, 554
 Goffman, E., 480
 Goldberg, D., 343
 Golde, P., 374
 Goldsen, J. M., 507
 Goldstein, M. C., 83, 84
 Gölge, Z. B., 91
 González, N. S., 375
 Gonzalez, P., 414
 Good, K., 348–49, 430
 Goode, E., 512
 Goodenough, W., 306–7, 542
 Goodman, L., 618
 Goodwin, C., 487
 Gooren, L. J. G., 46
 Gorden, R. L., 211, 217, 225
 Gottschalk, L. A., 515
 Gould, R. A., 445
 Gove, S., 352
 Graham, L. R., 422, 484
 Graham, M. A., 354–55
 Graves, T. D., 441, 667–68
 Gravlee, C. C., 31, 254
 Greenwald, A. G., 640
 Gregor, T., 614, 614

- Gregson, S., 93
 Grieco, E. M., 46
 Griffin, J. H., 437
 Grimshaw, J., 420
 Gross, A. E., 136
 Gross, D. R., 83, 84, 425, 427, 431
 Guba, E. G., 494
 Guest, G., 326
 Guilmet, G. M., 414
 Gujarati, D. N., 672
 Guttman, L., 324

 Hadaway, C. K., 91
 Hall, E. T., 441
 Halpin, M., 465, 466
 Hammel, E. A., 316
 Handwerker, W. P., 80, 86, 161, 353, 596,
 663, 678, 678–80, 679, 680
 Hanks, William, 478
 Hansen, A., 339–40
 Hansley, W. E., 284
 Harari, H., 136, 141
 Hargreaves, D. J., 94
 Harrington, J. P., 534
 Harris, M., 12, 85, 132–33
 Harrison, A., 341
 Harrison, E., 293–94
 Harrison, G. G., 447
 Harshbarger, C., 162, 601–2
 Hartman, D. A., 449
 Hartman, J. J., 163, 165
 Hartmann, D. P., 425
 Hatch, D., and M. Hatch, 152
 Hauser, R. M., 289
 Hayek, F. A. von, 20
 Hazan, A. R., 93
 Heath, D. B., 221
 Heberlein, T. A., 285
 Heckathorn, D. D., 193–94
 Hedblom, J., 163, 165
 Hedican, E. J., 542
 Helm, J., 449
 Hemmings, A. B., 2
 Henderson, J., 534

 Henley, N. M., 302–3
 Henry, G. T., 168
 Henry, J., 415, 417
 Herbert, S., 450
 Herdt, G., 343
 Hernández Jiménez, A., 466
 Herzfeld, M., 474–75
 Herzog, A. R., 283
 Hewett, P. C., 253, 260
 Hiatt, L. R., 83
 Hielscher, S., 89
 Hirschman, E. C., 510–11
 Hochstim, J. R., 260
 Hodge, L. G., 590–91
 Holbrook, A. L., 257
 Holland, D., 72
 Horn, W., 248
 Hornik, J., 131–32
 Horowitz, I. L., 74
 Houck, L. D., 414
 House, J. S., 285
 Houseman, M., 300
 Howell, N., 376, 377, 449
 Huberman, A. M., 388, 402, 403, 457,
 494, 495, 497
 Hughes, W. W., 446, 448
 Hume, L., 343
 Humphreys, L., 439
 Hunt, M. M., 108
 Hunt, P. S., 346
 Hursh-César, G., 268
 Hutchins, E., 481
 Hutt, S. J., and C. Hutt, 414
 Hyman, H. H., 217, 240
 Hymes, D., 475–76, 544
 Hymes, V., 476

 Ibeh, K. I. N., 258
 ICMR (Institute for Coastal and Marine
 Resources), 305
 Italie, G., 495
 Ives, E., 230
 Ivis, F. J., 269

 Jackson, B., 230
 Jacobson, L., 244

- Jacoby, H., 65
 James, J. E., 607
 Jamieson, P., 65
 Jargowsky, P. A., 52
 Jefferson, G., 487
 Jenkins, C., 602
 Jensen, P. M., 92
 Jobe, J. B., 249
 Johnson, A., 401, 427–28, 430, 434
 Johnson, J. C., 193, 305–6, 314
 Johnson, M., 482
 Johnson, R. K., 91, 297
 Johnstone, B. K. F., 264
 Johnstone, B. M., 126
 Jones, D. J., 358
 Jones, M., 340
 Jones, S. E., 441
 Joravsky, D., 6
 Jordan, B., 342, 424
 Jorgensen, D., 371
 Jussaume, R. A., Jr., 281, 284
- Kahn, R. L., 217, 239
 Kallan, J. E., 92
 Kane, E. W., 242
 Kaplan, D., 414
 Kaplowitz, M. D., 255
 Kasl, S. V., 93
 Kasser, T., 259
 Katsurada, E., 33
 Katz, D., 239
 Katz, K., 244
 Kearney, M. H., 500–503, 502
 Keesing, R. M., 472
 Keith, J., 340
 Kemper, R. V., 290, 350, 401
 Kempf, B. T., 259
 Kempton, W., 91, 481, 556, 615
 Kendall, C., 313, 542
 Kent, R. N., 425, 435
 Kenyatta, J., 471
 Ketterer, M. W., 607
 Kiecker, P., 263
 Kilbride, P. L., 369
- Killworth, P. D., 71, 214, 347, 401, 429
 Killworth, P. R. P., 343, 390
 Kimball, S. T., 397, 400
 Kinsey, A. C., 276
 Ki, Adam, 208–9
 Kleinman, A., 89
 Kluckhohn, K., 217
 Kneidinger, L. M., 414
 Knodel, J., 234, 236
 Koivusaari, R., 420
 Koocher, G. P., 440, 443
 Kortendick, O., 326–27, 341, 510
 Kraemer, H. C., 643
 Kreiss, L., 401, 403
 Krieger, L., 371
 Krippendorf, K., 510, 515
 Kroeber, A. L., 300, 449
 Kronenfeld, D. B., 300, 364
 Krosnick, J. A., 256, 269, 281
 Krueger, R. A., 233, 237, 238
 Kruskal, W., 618
 Kubey, R. W., 432
 Kulick, D., 376
 Kunitz, S. J., 644, 645, 646
 Kunovich, R. S., 121
 Kurasaki, K. S., 496, 497
- Laban Moogi Gwako, E., 353
 Labovitz, S., 638, 640, 646
 Ladd, E. C., 153
 La Flesche, F., 471
 Lakoff, G., 482
 Lame Deer, J., 472
 Landauer, T. K., 517
 La Pierre, R. T., 246
 Laraia, B. A., 93
 Larson, R., 156, 432
 Lastrucci, C. L., 5
 Latané, B., 131
 Laurent, C., 157–58
 Lave, C. A., 667–68
 Lazarsfeld, P. F., 232–33, 649
 Lea, K. L., 9
 Leach, E. R., 300, 355

- Leap, W., 376
 LeCompte, M. D., 355, 414
 Lees, S. H., 82
 Leggett, J. C., 241
 Lehman, F. K., 300
 Lehner, P. N., 414
 Leith, L. M., 121–22
 Lenski, G. E., 241
 Leslie, P. W., 248
 Levine, R. V., 396, 442, 444
 Levinson, D., 518–19
 Levy-Storms, L., 159
 Lewin, E., 376
 Lewis, O., 34
 Lieberman, D., 308
 Lieske, J., 676
 Lightcap, J. L., 79
 Likert, R., 327
 Lincoln, Y. S., 494
 Lindeman, M., 32–33
 Lininger, C. A., 217, 267–68, 268
 Livingstone, S., 693
 Locke, J., 11
 Loether, H. J., 618
 Lofland, L. H., 216, 220, 453, 503
 Loftin, C., 128
 Loftus, E. F., 249
 Loosveldt, G., 266
 Lowe, J. W. G., and E. D. Lowe, 449
 Lowie, R. H., 361
 Lowrey, T. M., 92
 Lucretius Carus, Titus, 7
 Lueptow, L. B., 242
 Lunt, P., 693
 Lurie, N., 472–73
 Lynd, R. S., and H. M. Lynd, 346

 Maercklein, G. H., 607
 Mahaffy, K. A., 191
 Malinowski, B., 344–45, 392
 Manzo, J., 488
 Marburger, W., 249
 Margolis, M., 454, 507–8
 Markowitz, F., 376

 Marquis, G. S., 436
 Marriott, B., 348
 Martin, J., 65
 Martínez, H., 531–34, 532, 533
 Matarazzo, J., 219
 Mathews, H. F., 50, 483–84
 Mathews, J. J., 471
 Mathewson, G. C., 131
 Matland, R. E., 65
 Max-Planck Institute, 16
 McCall, G., 414
 McCarrroll, J. E., 92
 McCaulay, L. J., 242
 McCormick, E. J., 274
 McDonald, D. D., 123–24
 McGue, M., 92
 McSpadden, L. A., 339–40
 McTavish, D. G., 618
 Mead, M., 360–61, 373
 Means, B., 249
 Meh, C. C., 414
 Mehta, R., 94
 Mensch, I. N., 415, 417
 Merton, R. K., 217, 232–33
 Metzger, D. G., 538
 Meyerhoff, B., 373
 Middlemist, R. D., 442–43
 Miles, M. B., 388, 402, 403, 404, 457, 494, 495, 497
 Milgram, S., 75, 76, 136, 137
 Mill, J. S., 13–14, 18
 Miller, A. S., 92
 Miller, K. W., 161
 Miller, L. L., 384–86
 Mills, J., 130–31
 Minton, A. P., 91
 Mitchell, R., 243
 Mizes, J. S., 284
 Momaday, N. S., 471–72
 Monárrez-Espino, J., 305
 Moore, R. P., 242
 Morchio, G., 223
 Morgan, D. L., 233, 237
 Morgan, L. H., 82, 300

- Morin, R., 281
 Morrison, M., and T. Morrison, 108
 Morse, J. M., 510
 Mulcahy, F. D., 450
 Mulcock, J., 343
 Murdock, G. P., 85, 400, 468, 470, 516, 518
 Murphy, G. P., 447
 Murray, S. O., 343
 Murray, T., 76
 Murtagh, M., 418–19
 Mwangi, E., 62, 668–69
- Nabokov, P., 472
 Nachman, S. R., 200
 Nakamura, T., 92
 Naré, C., 244
 Naroll, R., 350, 361, 519
 National Cancer Institute, 127
 Nave, A., 265
 Nederhof, A. J., 281
 Neihardt, J., 472
 Nelson, J. E., 263
 Nerlove, S. B., 309, 310, 310
 Newman, K. S., 504
 Niebel, B. W., 17, 414
 Nies, M. A., 340
 Nightingale, F., 18
 Nkwi, P. N., 235
 Nock, S. L., 291
 Nooney, J. G., 91
 North, A. C., 94
 Nyamongo, I. K., 496, 498–99
- Oberg, K., 380
 Oboler, R. S., 428, 430
 O'Brian, R., 434
 O'Brien, M., 509
 O'Brien, T. O., 253
 Ogburn, W. F., 20
 Ogilvie, D. M., 520
 Osgood, C. E., 336
 Oskenberg, L., 262
 Otnes, C., 92
- Otterbein, K., 516
 Oyuela-Caycedo, A., 165–66, 166, 167
- Paddock, J., 508
 Pagliari, C., 420
 Papadopoulos, F. C., 130
 Paredes, J. A., 83–84
 Partridge, W. T., 397, 400
 Patel, J. D., 127
 Pauli, J., 359
 Pausewang, S., 246
 Paxon, H., 2
 Payne, S. L., 271
 Pearson, J., 418
 Pelto, G. H., 352
 Pelto, P. J., 64
 Pelto, P. J., and G. H. Pelto, 196
 Perchonock, N., 314, 315
 Pericliev, V., 542
 Peters, D. L., 274
 Petersen, L. R., 91
 Peterson, L., 259
 Peterson, M., 126
 Peterson, R. A., 243
 Pfeiffer, J., 195
 Phillips, M., 440
 Piliavin, I. M., 136, 141
 Plattner, S., 461, 462
 Poggie, J. J., 201, 201, 668
 Pollock, P. H., 128
 Potter, P. B., 445
 Powdermaker, H., 372, 390
 Prebisch, R., 39
 Presser, S., 239, 250, 264, 269
 Price, D. H., 82, 350
 Proctor, S. P., 92
 Puccio, E. M., 607
- Quandt, S. A., 374
 Quételet, A., 12, 18
 Quinn, N., 482–83, 522
- Radin, P., 360, 466
 Ragin, C. C., 545
 Rappaport, R., 376

- Rashid, R. G., 121
 Rathje, W. L., 446–47, 448
 Reboussin, D., 278
 Reed, T. W., 217
 Reese, S. D., 241
 Reiss, A. J., Jr., 414
 Repp, A. C., 435
 Rhode, S., 239
 Ribeiro, R. P., 66
 Ricci, J. A., 435
 Richards, A. I., 293
 Richardson, J., 449
 Richardson, S. A., 210, 217
 Ricoeur, P., 23
 Rindfleisch, A., 94
 Ritenbaugh, C. K., 447
 Rivers, W. H. R., 300
 Robbins, M. C., 200–201
 Roberson, L., 341
 Roberts, J. M., 312, 415, 416
 Robinson, M. B., and C. E. Robinson, 61
 Robinson, W. S., 51, 239
 Rogoff, B., 431
 Rohner, R., 199, 465
 Romney, A. K., 203, 204, 205, 207–8,
 208, 209, 304, 308, 311, 313, 375
 Rose, M. G., 91, 478
 Rosenberg, L., 194–95
 Rosenberg, S. D., 520
 Rosenhan, D. L., 438
 Rosenthal, R., 244
 Ross, H. L., 134–35
 Ross, M. W., 254–55
 Rossi, P. H., 291
 Roth, P. L., 281
 Rothschild, R. F., 10
 Rotton, J., 64–65, 143–45, 442, 634, 634
 Rousseau, J.-J., 12
 Roy, P., 268
 Royal Anthropological Institute, 345–46
 Royce, A. P., 290, 350
 Rozeboom, W., 640
 Rubin, D. B., 244
 Rushforth, S., 542
 Rutledge, E. M., 620, 620
 Ryan, G. W., 494, 495, 531–34, 532, 533,
 556, 589, 616
 Sackett, R., 108
 Sacks, H., 487, 489, 490
 Sahlins, M., 108
 Salant, P., 280
 Salinas Pedraza, J., 467–68, 469–70,
 478–79, 479
 Salmon, M. H., 77
 Sanchez, M. E., 223
 Sandall, Roger, 199–200
 Sands, K., 471
 Sanjek, R., 389, 401
 Sankoff, G., 306
 Sarkodie, S., 437–38
 Sarton, G., 12
 SAUS, 16, 30, 57, 255
 Savala, R., 471
 Scaglione, R., 426–27, 430, 431
 Schaeffer, N. C., 239
 Schank, R. C., 480
 Schegloff, E., 487, 490
 Schensul, S. L., 305
 Scheper-Hughes, N., 342, 374
 Scherer, S. E., 441–42
 Schiller, F. C. S., 21
 Schlegel, A., 61, 85–86, 518
 Schmeets, H., 283
 Schober, M. F., 256
 Schuman, H., 219, 222, 239, 269
 Schutz, A., 24
 Schwarz, B., 607
 Schwarz, N., 55, 239, 247, 249, 338
 Schweizer, T., 545–48
Science, 214
 Scribner, S., 252
 Scudder, T., 290
 Sechrest, L., 440, 444
 Sexton, J. D., 467
 Shamdasani, P. N., 233
 Sharpe, R., 438
 Shats, M., 143–45

- Shaw, C. R., 346
 Shaw, S., 294, 294
 Sheatsley, P. B., 276
 Sheets, J. W., 172
 Shelley, G. A., 226, 398–99, 399, 407, 551
 Sherry, J. F., Jr., 343
 Sherzer, J., 477–78, 479, 484
 Shostak, M., 472
 Simmons, L. W., 472
 Sirken, M. G., 272
 Skinner, B. F., 4
 Skinner, D., 72
 Skruppy, M., 93
 Smith, C. J., 297
 Smith, L. D., 19
 Smith, M. L., 107, 252
 Smith, P. C., 341
 Smith, P. H., 495
 Smith, T. W., 250, 259, 271, 338
 Smits, J., 632–33, 633
 Snider, J. G., 336
 Soifer, V., 6
 Sokolovsky, J., 602
 Solomon, C., 342
 Sommerfeld, J., 89
 Sparrow, N., 188
 Spector, P. E., 328, 330
 Spradley, J. P., 211, 216, 220, 301, 537
 Sproull, L. S., 414
 Squire, P., 279
 Srinivas, M. N., 381
 Stearman, A. M., 621, 622
 Stein, J., 676
 Stemmer, N., 4
 Sterk, C. E., 372
 Stevens, S. S., 637–38
 Steward, J. H., 82–83
 Stewart, D. W., 233
 Stewart, M. A., 420
 Stimson, R. J., 217
 Stinson, L., 250
 Stockton, E., 401, 403
 Stone, P. J., 520
 Stouffer, S., 65
 Strathern, A., 472
 Strauss, A., 402–4, 492, 493, 499, 503
 Streib, G. F., 218
 Studlar, D. T., 65
 Stunkard, A., 414
 Sturtevant, W. C., 390
 Sudman, S., 55, 247, 249, 272, 282, 296
 Sugarman, J. R., 155–56
 Sugihara, Y., 33
 Sugita, E., 87, 358–59
 Sullivan, K. A., 414
 Sullivan, P., 361–62
 Sundvik, L., 32–33
 Swan, D. C., 449
 Sykes, R. E., 414, 442
 Synodinos, N. E., 281
 Szalai, A., 291

 Tannen, D., 491
 Taylor, H., 262
 Taylor, S. J., 383, 387
 Tedlock, D., 476–77
 Tesch, R., 510
 Tewksbury, R., 94
 Thiemann, S., 643
 Thomas, J. S., 670–72, 671
 Thurman, Q., 285
 Tinsley, H. E., 187
 Tooker, E., 109
 Toth, J., 243
 Toulmin, S. E., 9
 Tourangeau, R., 259, 338
 Tremblay, M., 200
 Trent, R., 291
 Trost, J. E., 510
 Trotter, R. T., II, 303
 Truex, G., 407–8
 Turley, S. K., 281
 Turnbull, C., 215, 371, 375
 Twenge, J. M., 44, 45

 Uchendu, V., 471
 University of Michigan, 267
 Urban, G., 484

- Vadez, V., 55, 93
 Valdez-Perez, R. E., 542
 van der Geest, S., 438
 Van Maanen, J., 457, 457, 458
 Vannatta, R. A., 92
 Vasquez, L. A., 92
 Vaughn, S., 233
 Vieco Albarraçín, J. J., 165–66, 166, 167
 Vieth, A. Z., 91
 Vincke, J., 338
 Voltaire, 11

 Wagley, C., 379, 382
 Wainer, H., 640
 Waitzkin, H., 485–86
 Wallace, A. F. C., 365
 Wallerstein, E. M., 65
Wall Street Journal, 424
 Wanous, J. P., 341
 Wapner, S., 93
 Ward, T. W., 290
 Ward, V. M., 236
 Warner, W. L., 346
 Warren, C. A. B., 374
 Warriner, K., 285
 Warwick, D. P., 217, 267–68, 268
 Watson, O. M., 441
 Wax, R., 374, 379
 Weatherford, J. M., 440
 Webb, B. P., 346
 Webb, E. J., 445
 Webb, S., and B. P. Webb, 18
 Weeks, M. F., 242
 Weinberger, M., 188
 Weisner, T., 193, 495
 Weller, S. C., 129–30, 207–8, 208, 209,
 313, 317, 688, 688
 Wells, G. L., 250
 Werblow, J. A., 120
 Werner, D., 327
 Werner, O., 277, 314, 315
 West, S. G., 142
 White, D. R., 85, 300, 516, 518
 White, K. R., 108
 White, L. A., 82, 470

 Whitehead, T. L., 374
 Whiting, B. W., and J. W. M. Whiting,
 422–23
 Whiting, J. W. M., 343, 517
 WHO, 352
 Whyte, W. F., 197–98, 217, 222, 369
 Wiederman, M., 512
 Wiesel-Lev, R., 192
 Wilke, J. R., 26
 Will, J. A., 292
 Willett, W. C., 607
 Williams, B., 125, 125, 295
 Williams, G. E., 538
 Williams, H. A., 348
 Williams, T., 358
 Willimack, D. K., 281
 Willis, F. N., 512
 Willms, D. G., 494
 Willson, M., 376
 Wilson, D., 33
 Wilson, M., 79
 Wilson, R. P., 234–35
 Windelband, W., 81
 Withey, S. B., 340
 Wittfogel, K. A., 82
 Wolcott, H., 224, 228, 369, 468, 469, 470
 Wolf, D. R., 350, 357
 Wood, C. S., 694
 Wood, D. D., 425
 Woodbury, A., 479
 Worth, S., 467
 Wuthnow, R., 43

 Yamada, S., 281, 284
 Yammarino, F. J., 280
 Young, J. C., 305, 522, 527–30, 528, 529,
 530
 Yu, X., 349

 Zehner, R. B., 240
 Zhao, S., 264
 Zinner, D., 414
 Zipp, J. F., 243
 Zorbaugh, H. W., 346

About the Author

H. Russell Bernard (Ph.D., University of Illinois, 1968) is professor of anthropology at the University of Florida. He has also held faculty positions at Washington State University and West Virginia University and has taught or done research at the University of Athens, the University of Cologne, the National Museum of Ethnology (Osaka), the University of Michigan, and the Scripps Institution of Oceanography. Dr. Bernard works with indigenous people to develop publishing outlets for works in previously nonwritten languages. He also does research in social network analysis, particularly on the problem of estimating the size of uncountable populations.

His book (with Jesús Salinas Pedraza) *Native Ethnography: A Mexican Indian Describes His Culture* (Sage, 1989) won special mention in the 1990 Chicago Folklore Prize. Other books include *Technology and Social Change* (Waveland, 1983, 2d ed., edited with Pertti Pelto), three editions of *Research Methods in Anthropology* (Sage, 1988; AltaMira, 1994, 2000), *Data Analysis with MYSTAT* (McGraw-Hill, 1995, with W. Penn Handwerker), *Handbook of Methods in Cultural Anthropology* (AltaMira, 1998), and *Social Research Methods* (Sage, 2000).

Dr. Bernard received the 2003 Franz Boas Award from the American Anthropological Association for his service to the discipline. He was editor of the journals *Human Organization* (1976–1981) and *American Anthropologist* (1981–1989), and he founded and currently edits the journal *Field Methods*. He has worked since 1987 in NSF-supported research methods training programs. He is one of the founders of the International Social Network Conference and serves on Board of Directors of the Human Relations Area Files.