Theory Construction and Model-Building Skills

A Practical Guide for Social Scientists

James Jaccard | Jacob Jacoby

Theory Construction and Model-Building Skills

Methodology in the Social Sciences

David A. Kenny, Founding Editor

Todd D. Little, Series Editor

This series provides applied researchers and students with analysis and research design books that emphasize the use of methods to answer research questions. Rather than emphasizing statistical theory, each volume in the series illustrates when a technique should (and should not) be used and how the output from available software programs should (and should not) be interpreted. Common pitfalls as well as areas of further development are clearly articulated.

SPECTRAL ANALYSIS OF TIME-SERIES DATA
Rebecca M. Warner

A PRIMER ON REGRESSION ARTIFACTS
Donald T. Campbell and David A. Kenny

REGRESSION ANALYSIS FOR CATEGORICAL MODERATORS
Herman Aguinis

HOW TO CONDUCT BEHAVIORAL RESEARCH OVER THE INTERNET:
A BEGINNER'S GUIDE TO HTML AND CGI/PERL
R. Chris Fraley

PRINCIPLES AND PRACTICE OF STRUCTURAL EQUATION MODELING, SECOND EDITION
Rex B. Kline

CONFIRMATORY FACTOR ANALYSIS FOR APPLIED RESEARCH
Timothy A. Brown

DYADIC DATA ANALYSIS
David A. Kenny, Deborah A. Kashy, and William L. Cook

MISSING DATA: A GENTLE INTRODUCTION
Patrick E. McKnight, Katherine M. McKnight, Souraya Sidani, and Aurelio José Figueredo

MULTILEVEL ANALYSIS FOR APPLIED RESEARCH: IT'S JUST REGRESSION!
Robert Bickel

THE THEORY AND PRACTICE OF ITEM RESPONSE THEORY
R. J. de Ayala

THEORY CONSTRUCTION AND MODEL-BUILDING SKILLS:
A PRACTICAL GUIDE FOR SOCIAL SCIENTISTS
James Jaccard and Jacob Jacoby

Theory Construction and Model-Building Skills

A Practical Guide for Social Scientists

James Jaccard | Jacob Jacoby

Series Editor's Note by David A. Kenny

THE GUILFORD PRESS
New York London

A Division of Guilford Publications, Inc.
72 Spring Street, New York, NY 10012
www.guilford.com

Printed in the United States of America

This book is printed on acid-free paper.

Last digit is print number: 9 8 7 6 5 4 3 2 1

Library of Congress Cataloging-in-Publication Data

Jaccard, James.
Theory construction and model-building skills : a practical guide for social scientists / by James Jaccard and Jacob Jacoby.
p. cm. — (Methodology in the social sciences)
Includes bibliographical references and index.
ISBN 978-1-60623-339-9 (pbk.) — ISBN 978-1-60623-340-5 (hbk.)
1. Social sciences—Research—Methodology. 2. Theory (Philosophy)
I. Jacoby, Jacob. II. Title.
H62.J29 2010
300.72—dc22

2009040114

For Marty Fishbein—a brilliant and inspiring theorist and mentor

—JAMES JACCARD

For Renee and Dana—in appreciation for the balance and joy they create

—JACOB JACOBY

Series Editor's Note

For the first time in my life, I feel a bit like a famous athlete. That athlete is the quarterback Bret Favre. As you may already know, Bret Favre was a Super Bowl-winning quarterback for many years with the Green Bay Packers. Two years ago, he tearfully retired from football, but not long after that he came out of retirement to be the quarterback for the New York Jets. I too retired as Series Editor of Methodology in the Social Sciences for The Guilford Press. However, when I saw this project I gladly came out of retirement to serve as the coeditor with Todd Little. I am happy to report that this project appears to be much more successful than Favre's controversial year with the Jets. This book promises to be a valuable resource to social and behavioral sciences students and faculty.

We all know that when we write a journal article, a dissertation, or grant proposal, we need to have a theory. For many of us, this task is not so straightforward. I know from over 40 years of reading student papers and reviewing articles that many researchers have a very difficult time framing their research theoretically. They do not know where to start. I have had a difficult time trying to find a way to teach my students to think in terms of theory. Most of the resources on theory building that I would send my students to were dreadfully boring and, even worse, not very practical for the task that faced them. Finally, now I have something that can help them.

The book gives clear suggestions to the reader on how to come up with a theory. The two authors, who each have illustrious careers in two very different areas, one in attitudes and the other in consumer decision making, have combined to provide a readable and practical discussion of theory construction. The book provides a useful source for helping researchers come up with ideas for research and for fine-tuning the resultant theories that emerge from such thinking. Thankfully, they choose not to provide an abstract, formal guide to theory construction. Rather, they provide practical help illustrating cognitive heuristics and tricks of the trade that they have used in their careers.

I find it particularly interesting that the authors use ideas from a wide range of research methods to assist the reader in theory construction. They are equally comfortable discussing randomized experiments, mathematical modeling, simulations, causal modeling, and qualitative methods in this endeavor. The authors adopt very much of a "mixed methods" approach to theory development. Researchers should try out several different approaches and eventually there will be one method that suits the researchers' style and content. Although it may be impossible to teach people to be creative, this book comes the closest to doing so. It certainly challenges the reader to become more creative.

In sum, this book definitely scores a touchdown. In fact, it scores a good number of touchdowns and I feel it deserves to win the equivalent of publishing's Super Bowl!

David A. Kenny

Preface

Theory construction is at the heart of the scientific process. The strategies that social scientists use to generate and develop ideas are important to understand and foster in young academics and investigators as they prepare for a research-oriented career. Although books have been written about theory construction, there are surprisingly few books on the topic that tackle the problem of teaching students and young professionals, in a practical and concrete way, how to theorize. Students, especially graduate students, take one or more courses on research methods and data analysis, but few experience more than a lecture or two, or read a chapter or two, on theory construction. It is no wonder that students often are intimidated by the prospect of constructing theories.

This book provides young scientists with tools to assist them in the practical aspects of theory construction. It is not an academic discussion of theory construction or the philosophy of science, and we do not delve too deeply into the vast literature on these topics. Rather, we take a more informal journey through the cognitive heuristics, tricks of the trade, and ways of thinking that we have found to be useful in developing theories—essentially, conceptualizations—that can advance knowledge in the social sciences. By taking this journey, we hope to stimulate the thinking and creative processes of readers so that they might think about phenomena in new and different ways, perhaps leading to insights that might not otherwise have resulted. The intent of this book is to provide a practical, hands-on, systematic approach to developing theories and fostering scientific creativity in the conceptual domain. Relative to the vast majority of books on theory construction, this book is unique in its focus on the nuts and bolts of building a theory rather than on an analysis of broad-based systems of thought.

Science is about understanding nature and the reasons for things. It is one of humanity's greatest ongoing adventures. This book is intended to help propel two types of readers along this exciting journey. First, the book is written for graduate and advanced undergraduate students interested in pursuing careers as researchers in the social sciences, as well as for newly minted PhD social scientists. Second, the book also should benefit those who desire to pursue a professional career in the social sciences, but who do not plan on becoming researchers. It will help them understand and evaluate the theories they read in professional journals and identify gaps in those theories. It will help them think about theories from different vantage points.

The book can be used in many different disciplines. We draw on examples from the fields of anthropology, business, communications, education, economics, health, marketing, organizational studies, political science, psychology, social work, and sociology, to name a few. Some instructors may prefer more detailed examples in their particular field of study, but we believe that using examples from multiple disciplines helps students appreciate the commonalities and value of multidisciplinary perspectives.

We have used drafts of the book as both a stand-alone text in a course on theory construction as well as one of several texts in graduate courses on research and research methodology. In terms of the latter approach, almost all traditional research methods books include a section or chapter on the nature of theory and/or theory construction. However, the treatment of theory construction usually is brief and of limited practical value. The present book is intended to provide the instructor with a useful source for helping students come up with ideas for research and for fine-tuning the resultant theories that emerge from such thinking. It provides more detail and more practical knowledge than what is typical of chapters in books on research methodology. The social psychologist William McGuire often lamented about how research training with graduate students focuses at least 90% on teaching methods to test ideas, but no more than 10% on how to get those ideas in the first place. Despite this difference in emphasis, the process of theory development is fundamental to successful scientific research. Indeed, many would say that there can be no theory testing without prior theory. An objective of this book is to move toward a much-needed balance in the emphases given to theory construction and theory testing.

In our research methods courses, we assign this book to be read during the first 2–3 weeks of classes. We allow the book to stand on its own as a teaching device, and we spend 1 or 2 weeks of lectures/discussion on the material. Obviously, not all of the material can be covered in these sessions, so we select material to address based on the needs of the students. Part of the in-class coverage includes providing students with a verbal "road map" to the book and an overview of each chapter. We stress that students must read the entire book, and we test them on the assigned material with essay and short-answer questions from a small sampling of the book. For some student cohorts, we do not assign certain chapters, depending on the students' areas of emphasis. For example, students primarily oriented toward qualitative research might not be assigned the chapters on mathematical modeling and/or simulations. However, we omit chapters with reluctance, as our goal is to expand the theoretical toolbox of students.

In our courses on theory construction, we use the book as our main text and cover each chapter during lecture/discussion sessions. We ask students to read one or more chapters the week before class, then discuss those chapters the following week. After a few weeks of class, we start to assign empirical articles in journals that the students are to read. We ask one student to orally summarize in class the theory being espoused in the assigned article and to critique it. After the student has done so, we open discussion to all class members, who add further commentary or analysis. It is not uncommon for students to extract different interpretations and representations of the theory, and it is instructive when the class then sets about the task of resolving the discrepancies. Finally, we provide feedback relative to our own analysis and critique of the theory. The level of analysis becomes increasingly sophisticated as students learn more of the material covered in the book.

A major assignment in the theory construction class is for each student to construct his or her own theory about a phenomenon of interest to him or her. We help students select a topic in the first weeks of class and then direct them to the relevant literatures to read and consider. Midway through the course, students present their initial theory, and we help them refine it, expand it, and finalize it over the rest of the course. We do this as part of an in-class exercise for each student, so that all students in the class can observe the creativity and process of refinement for each project. Indeed, we encourage other class members to contribute to and participate in the developmental process for each project. This repetition across projects helps students better internalize the principles of theory construction.

The book has several pedagogical features that enhance its use as a textbook and as a source of learning. First, each chapter includes a section on suggested readings with commentary, where we direct the reader to key references for further study on the topics covered in the chapter. Second, each chapter has a list of key terms that highlights the most important jargon and terminology. Third, each chapter has a set of exercises that encourages the reader to think about the material that was presented in the chapter. We include exercises to reinforce concepts and exercises to apply the concepts to problems of interest. Finally, each chapter has a highlighted box that covers an interesting topic that applies the concepts covered in the chapter or that shows important uses of them.

We bring a wide range of personal experience in diverse research settings to this project, including laboratory studies, small-scale and large-scale surveys, simulations, and studies in naturalistic settings. Our research has attempted to advance basic theory as well as solve applied problems. We have worked with interdisciplinary teams, and we have "gone it alone." We have interacted extensively with advertising researchers, anthropologists, biologists, communication theorists, demographers, economists, educationalists, epidemiologists, health researchers, historians, legal scholars, marketers, political scientists, social workers, and sociologists, and we appreciate the perspectives of these diverse disciplines. We have brought these perspectives to bear in this book. Nevertheless, we recognize how limited and constraining our experience has been relative to the vast array of topics and perspectives that have been pursued in the social sciences. As such, this book is best viewed as a somewhat limited personal account of our own perspectives on the process of theory construction.

We appreciate that this book's descriptions of theory construction may have more to say to "quantitatively oriented" than to "qualitatively oriented" behavioral scientists and to "variable-oriented" as opposed to "process-oriented" theorizing. In part, this is because variable-oriented theories have been more dominant in the behavioral sciences, at least in some disciplines (with the most notable exception being anthropology). That said, we have tried to give both perspectives their due, and we believe that both are well represented in these pages. In our opinion, good theories incorporate both process- and variable-oriented perspectives, so it is important to consider both approaches without preconceived biases for or against one or the other. We think that scientists should have a broad theoretical toolbox from which to draw. Perhaps it is sometimes useful to think about phenomena in terms of "variables," and perhaps it is sometimes useful to think about phenomena in terms of "processes." This book encourages thinking from multiple perspectives rather than from one school of thought. In this sense, readers committed to a given method of thinking will need to keep an open mind.

When we have used the book with more qualitative, process-oriented students, we omit the chapters on mathematical modeling and simulations, as noted, but all other chapters are relevant. To be sure, we note to the students that some of the chapters are more "variable oriented" (e.g., the chapters on clarifying relationships using thought experiments and causal modeling), but we encourage the students to approach these chapters with open minds that might provide them with a unique way of thinking about matters as they adopt the more familiar grounded/emergent theory approach of Chapter 10. We alter the order in which we ask qualitatively oriented students to read chapters, starting with Chapters 1–5 as introductions, then Chapter 10 on grounded theory, then Chapter 6 on clarifying relationships, and then Chapter 7 (Causal Models), Chapter 11 (Historically Influential Systems of Thought), Chapter 12 (Reading and Writing about Theories) and Chapter 13 (Epilogue).

ACKNOWLEDGMENTS

Many people have contributed to the development of this book, in the form of inspiration, critical feedback, and/or in the tangible aspects of bringing the book to fruition. First, we would like to thank students and colleagues who provided us with feedback on earlier versions of the book, including Hart Blanton, David Brinberg, Jon Hobbs, Peter Schultz, Wendy Silverman, Heather Taylor, Daniel Whitman, and Dan Wright. We also received many helpful comments and suggestions from the reviewers: Elif Andac, Department of Sociology, University of Kansas; Robert Bickel, Department of Education, Marshall University; Thomas A. Cornille, Department of Family and Child Sciences, Florida State University; Joseph Maxwell, Graduate School of Education, George Mason University; Richard Tardanico, Department of Sociology and Anthropology, Florida International University; and Rob Turrisi, Department of Biobehavioral Health, Pennsylvania State University. Liliana Goldin shared invaluable perspectives and wisdom both in the form of feedback and acting as a sounding board for many ideas. David Kenny provided insightful and useful comments at many different levels, always in a constructive way, proving to be adept at helping us think through issues. His focus on the "big picture" also proved invaluable. At The Guilford Press, C. Deborah Laughton was outstanding in her role as editor for this project. Every task she tackled was handled in a professional and thoughtful manner that improved the book immensely. We are

grateful to her for her input and support. Although all of these individuals contributed significantly to the book, we alone are responsible for any of its shortcomings.

We have been taught and inspired by so many professors and colleagues over the years that we dare not list them all here, but we express sincere gratitude for their efforts, both as teachers and scientists. I (Jaccard), however, would like to extend a special note of appreciation to Marty Fishbein, to whom this book is dedicated. Marty has been a one-of-a-kind role model as a scientist, teacher, professor, and mentor and has touched the lives of his students in so many positive ways. This book never would have come to be without his support, guidance, inspiration, and insights over the years. We also want to thank our many students for their inquisitive minds, for allowing us the opportunity to watch them grow into first-rate scientists, and for contributing to our own intellectual development in ways they probably cannot begin to realize. It has been our good fortune to learn so much from our students.

Jack Jacoby's spouse, Renee, was, as always, supportive, patient, and understanding throughout the many hours the project absorbed. Jim Jaccard's spouse, Liliana, was none of these. As a creative social scientist in her own right, she cajoled, encouraged, criticized, and ultimately shaped core parts of this project, which has been typical of past collaborations with him, both formal and informal alike. Their best collaborative effort, however, has been their daughter, Sarita, who inspired ideas in this book in special ways. It goes without saying that this book never would have seen the light of day without Renee, Liliana, and Sarita being a central part of our respective lives.

Brief Contents

Contents

Part I

BASIC CONCEPTS

1

Introduction

The central role of theory in the social sciences is disputed by few. Scientists formulate theories, test theories, accept theories, reject theories, modify theories, and use theories as guides to understanding and predicting events in the world about them. A great deal has been written about the nature and role of theory in the social sciences. These writings have spanned numerous disciplines, including anthropology, economics, history, philosophy, political science, psychology, sociology, and social work, to name but a few. This literature has described, among other things, broad frameworks for classifying types of theories, the evolution of theories over time, the lives and scientific strategies of great scientific theorists, and general issues in the philosophy of science. Although this literature is insightful, much less has been written to provide social scientists with practical guidelines for constructing theories as they go about the business of doing their science. Most students are intimidated by the prospect of constructing their own theories about a phenomenon. Theory construction is viewed as a mysterious process that somehow "happens" and is beyond the scope and training of a young scientist trying to find his or her way in the field. Whereas most graduate programs in the social sciences require multiple courses in research methodology to ensure that students become equipped with the tools to test theories empirically, the same cannot be said for theory construction. In contrast to focusing on methods for testing theory, this work focuses on methods for generating theory.

The fundamental objective of this book is to provide students and young scientists with tools to assist them in the practical process of constructing theories. It does so via describing in some detail the strategies, heuristics, and approaches to thinking about problems that we have found to be useful over the more than 70 collective years that we have been doing social science research. This book is not an academic discussion of the literature on theory construction or the philosophy of science. We do not delve too deeply into the vast literature on these topics. Rather, we take a more practical journey through the cognitive heuristics, tricks of the trade, and ways of thinking that we have found to be useful in developing theories.

ORGANIZATION OF THE BOOK

The book is organized into four parts. Part I presents the basic concepts that form the backdrop for later chapters. In these early chapters we consider the nature of science and what it means to understand something. We develop the notion of *concepts* and highlight the central role of concepts in theories. We lay the foundations for communicating to others the concepts in one's theory and then describe what separates science from other ways of knowing.

With this as background, we turn to developing core strategies for constructing a theory, the topic of Part II. In Chapter 4 we focus first on strategies for generating ideas and for stimulating creative thinking. Once you have a set of rough ideas, they need to be refined and focused to meet the criteria of a rigorous scientific theory. Chapter 5 describes strategies for thinking through your constructs and discusses how to develop clear and communicable conceptual definitions of them. We provide numerous strategies for making fuzzy constructs more precise and not overly abstract. Chapter 6 focuses on relationships between variables and develops strategies for making explicit the relationships you posit between variables. We show how to derive theoretical propositions based on a careful analysis of relationships.

Part III considers different frameworks for theory generation. Chapter 7 considers one of the most dominant approaches to theory construction in the social sciences: the framework of causal thinking. This approach elaborates the causes and consequences of different phenomena and views the identification of causal linkages as a central goal of science. Chapter 8 describes strategies for building mathematical models of different phenomena. Our intent here is to make clear the sometimes seemingly mysterious ways in which mathematics and social science theorizing interface. The chapter includes a brief introduction to chaos theory and catastrophe theory as examples of mathematical models. Chapter 9 describes the potential that simulations—in particular, the development of simulations—have for theory construction. Chapter 10 develops grounded and emergent approaches to theory construction, which tend to rely on qualitative methods to identify constructs and relationships on which to focus a theory. The spirit of these approaches is that the theory emerges from the data. Some social scientists might argue that this chapter belongs in the previous section, where one initially identifies constructs and relationships to include in a theory. As we emphasize throughout this book, theory construction is not a set process, and a case like this could be made for almost every chapter in the current section. This book provides you with key ingredients for constructing a theory. How you choose to mix those ingredients to form your theoretical recipe depends on your predilections and the domains that you are studying. Chapter 11 summarizes 12 broad-based theoretical frameworks that may help in the idea-generation process. These frameworks include materialism, structuralism, functionalism, symbolic interactionism, evolutionary perspectives, postmodernism, neural networks, systems theory, stage theories, reinforcement theories, humanism, and multilevel modeling.

In the final section we discuss strategies for reading journal articles and scientific reports so as to make explicit the theories that the authors describe and subject to empirical evaluation. We also discuss strategies for presenting theories in different kinds of reports. We close with an Epilogue that comments on the theory construction process in light of the material covered in previous chapters and that addresses some odds and ends that did not fit well into the other chapters.

THEORIES AND SETTINGS

This book is written primarily for students and professionals interested in pursuing a career as a researcher in the social sciences. It is intended to provide you with concrete strategies for building upon existing theories and constructing your own theories. Theorizing does not occur in a vacuum. It occurs in the context of individuals pursuing a career in some professional setting, usually an academic setting. At times, we describe how the setting in which you work impacts the way in which you theorize and the kinds of questions you ask. We also discuss strategies for dealing with the constraints you face as a result of these settings.

Considering how to develop theory flows quite naturally into considering the empirics involved in testing theory and the important interplay between the two. Yet to have included more than minor considerations of empirics not only would have generated a work of monumental size, but also would have diluted what we view as a necessary focus on theory construction, a topic we believe is seriously shortchanged in the education of social scientists. It is for this reason that this work contains fewer empirical illustrations than some might desire or expect.

For us, constructing theory is one of the most rewarding aspects of doing science—on par with the excitement associated with empirically testing and finding support for theory. In all honesty, we probably are more captivated by research that questions the theories we have posited because of the ensuing call to "put on our detective hats" to figure out why we were wrong. This invariably demands that we approach the problem from a new conceptual angle. This book describes some (but not all) of the types of detective hats that we have put on over the years. We hope that we can help to start you down the path of a more enriched and productive approach to the construction of theories as you fashion your own strategies and set of detective hats for thinking about phenomena and solving problems.

2

The Nature of Understanding

Reality is merely an illusion, albeit a very persistent one.

—Albert Einstein

The whole of science is nothing more than an extension of everyday thinking.

—Albert Einstein

Despite the fact that science has been practiced for thousands of years and countless books have been written on the subject, many people still consider it mysterious and forbidding. Perhaps the reason for this is because they view science as something fundamentally different from anything they normally do. Actually, this is not the case. The essence of science is something we all do, which is to try to understand ourselves and the world around us. Scientific research is a process that is designed to extend our understandings and to determine if they are correct or useful. The basic difference between everyday thinking, on the one hand, and science and scientific research, on the other, is that the latter strives to operate according to a more rigorous set of rules. Because science and the process of scientific research can be viewed as extensions of everyday thinking, we find them easiest to explain if we begin by considering how an individual tries to make sense of, and cope with, his or her world.

The present chapter explores the nature of understanding, relying on informal and everyday examples of human thought to draw parallels to scientific conceptions of understanding. In doing so, we build on Albert Einstein's assertion that "the whole of science is nothing more than an extension of everyday thinking." We begin by describing the different ways in which social scientists think about reality, considering the perspectives of realism, social constructionism, critical realism, and hypothetical realism. Next we address the building blocks of human understanding, namely, concepts and conceptual systems that relate one concept to another. Given that a scientist has evolved a conceptual system to address an issue, he or she then must communicate that system to other scientists. We conclude the chapter by briefly considering the nature of communication so as to set the stage for future chapters on how to derive precise conceptual definitions in theory construction.

THE NATURE OF REALITY

The process of understanding our world and making sense of reality is central to our waking lives. Accordingly, let us be more specific about what we mean by *reality* and *understanding*. Much philosophical thought has been devoted to the question of the nature of reality, and there is controversy among scientists about whether a single objective reality could ever be shown to exist.[1] According to the traditional perspective, termed *realism*, reality exists independent of any human presence. There is an external world comprised of objects that follow a myriad of natural facts and laws. It is up to us to *discover* these facts and laws. Using this perspective, science has evolved and prospered as an approach for gaining knowledge that mirrors the presumed actualities of the real world.

In contrast to realism, the *social constructionist* perspective holds that reality is a construction of the human mind, that this construction is tied to a particular time and social context, and that what is considered reality changes as the social context changes. In its most extreme form, constructionism maintains that there is no reality and there are no facts until these are conceptualized and shared by some number of people. A more moderate position holds that, though there is an external reality independent of humankind, we can never know its units and true laws—or even if it has units and true laws. All we can know is our interpretation or construction of these experiences. Since the same experiences are open to many interpretations, any or all of which may be correct, the correctness of an interpretation depends upon the purposes of those doing the interpreting. Thus, as Scarr (1985) notes:

> We do not discover scientific facts; we *invent* them. Their usefulness to us depends both on shared perceptions of the facts (consensual validation) and whether they work for various purposes, some practical and some theoretical. (p. 499; emphasis added)

As a simple example that drives the point home, we frequently draw a set of parallel lines on a blackboard and ask our students to describe what they see. Some reply "a road," others say "two lines." When an attempt is made to focus their thinking by saying "Hint: It's a number," three principal interpretations emerge: "Arabic number 11, Roman numeral II, or binary number three." Each of these responses represents a different, but potentially accurate, reconstruction of the same objective reality that reflects that individual's mental perspective. This point is fundamental. Even if the existence

[1]The discussion that follows is a simplified and not necessarily universally shared perspective on realism, social constructionism, and hypothetical realism. Philosophers and social scientists use these terms in different ways.

of a single objective, external reality could be assumed, the way in which this reality is interpreted can vary within the individual over time, across individuals, and can be heavily influenced by context (e.g., someone immersed in the study of Roman antiquity is more likely to interpret two parallel lines as representing Roman numeral II). Every individual develops his or her own reality, so that a number of different realities may be constructed out of the same set of "objective" facts. As is increasingly being recognized, it is possible for more than one of these different realities to be correct and useful.

The social constructionist perspective has implications for the way in which science is viewed (see, e.g., Gergen, 1985; Gergen & Gergen, 2003). The principal implications of the social constructionist perspective do not affect so much the way in which scientific empiricism is practiced, but rather the way in which the conceptualizations and the outcomes of the assessment process are interpreted. According to the realism perspective, conceptual systems and theories are created so as to precisely mirror an existing reality. The outcomes of the assessment process can be taken as direct representations of that reality, and it is possible to make claims regarding ultimate truths. By contrast, although the social constructionist perspective usually involves the researcher doing virtually the same things as are done in a realism perspective, the recognition that there exist multiple possible realities orients the researcher toward interpretations that reject more absolute perspectives on mapping out a single, existing reality:

> The admission that reality is a construction of the human mind does not deny the . . . value of the construction. Indeed, we get around in the world and invent knowledge that is admirably useful. But the claim that science and reality are human constructions denies that there is any one set of facts that is absolute and real. Instead, it asserts that there are many sets of "facts" that arise from different theory-guided perceptions. (Scarr, 1985, p. 501)

The social constructionist perspective can be discomforting because it makes us less certain of what we do and what we think we know. "How can we know what is right if there is no right?" (Scarr, 1985, pp. 511–512). On the other hand, this perspective enables us to more clearly recognize that any given conceptualization, and the facts that are given meaning by that conceptualization, is a function of the sociocultural time and space in which they occur.

A middle ground relative to these somewhat conflicting perspectives has been articulated by Blumer (1969). According to this view, *reality* is indeed seen through human conceptions of it. But the empirical world also "talks back" to our conceptions in the sense of challenging, resisting, and failing to bend to them. If a knife is plunged into someone's heart, certain ramifications follow. The ways in which these ramifications are construed and interpreted may vary from one conceptual scheme to another. But the environment has spoken. It is this inflexible character of the world about us that calls for, and justifies, empirical science. Science seeks to develop conceptions that can successfully accommodate the obdurate character of the empirical world. Blumer's view roughly maps onto a philosophy of science known as *critical realism*, though alternative formulations of it are many (e.g., Manicas, 2006; Sayer, 1992).

Another influential perspective on the debate is that even if it cannot be proven that reality exists, it is useful to assume that it does. This approach has been termed *hypothetical realism*. Here the concept of reality is a heuristic device—something that helps us organize our thoughts and think about matters so as to accomplish certain goals and objectives. Strictly speaking, reality may or may not exist, but we approach the world and our attempts to understand the world as if it does. In doing so, we may be able to accomplish a wide range of goals, but we also may be constrained in our thinking, accordingly. Hypothetical realism derives from a broader approach to epistemology, called *pragmatism*. The approach is reflected in the work of the philosopher C. I. Lewis (1929), who argued that science does not provide a copy of reality but must work with conceptual systems that are chosen for pragmatic reasons so as to aid scientific inquiry. Assuming a hypothetical reality is one such aid.

In sum, whereas realism embraces a view that external reality exists and the goal of science is to discover the laws that govern that reality, constructionism emphasizes that reality is a construction of the human mind that is tied to a particular time and social context. There are many gradations of these viewpoints, such as the position advocated by Blumer (1969), which emphasizes that reality is seen through human conceptions of it but that there is an empirical world that "talks back" to our conceptions; and hypothetical realism, which recognizes that one may not be able to prove that reality exists, but nevertheless approaches science with a working assumption that it does.

The broader literature on the philosophy of science explores a myriad of perspectives on how scientists (and laypeople) think about reality. This literature presents more nuanced perspectives than what we present here, and interested readers are encouraged to pursue this literature (see the suggested readings at the end of the chapter).

How Reality Is Experienced

Assuming for the moment that reality exists, how is it experienced by the individual? Most would agree that we experience the world around us as a complex, dynamic flow of unique and unrepeatable phenomena and events. Further, most of these phenomena and events—from those occurring deep in intergalactic space to those occurring in the micromolecular structure of this book—are not directly observable. No wonder, then, that attempting to understand our world can be a difficult process.

Reality appears complex. Whatever else it is or may be, the world—especially the external world—that we experience is complex. Consider a lecture hall filled with 200 students. Forget about the world beyond our immediate view; to describe, in precise detail, the sizes, shapes, colors (of clothing, objects, etc.), relationships, psychological components, and sociological components of that lecture hall at one instant in time could take months, years, or perhaps even lifetimes.

Reality appears dynamic. Moreover, things never stay the same. The world at any given instant is different from the world at the very next instant. From the tiniest particles that constitute physical matter to the largest galaxies, things are always in motion. The cells of living organisms are always growing or decaying, and the impulses in the neuronal system are always at work. So even if we were able to describe in infinite detail our hypothetical lecture hall filled with students, once one or more students moved, we would have a set of different relationships and, hence, a different reality.

Reality appears unique. Because of this dynamic quality, the universe at any given instant—and everything in it—is never the same as the universe at any other instant, either previous or subsequent. The water that flows at one particular instant or during any given day down the rivers of New Hampshire, the raindrops that fall on a particular evening in Houston, the expense account dinner that was eaten in Paris—all are unique and can never be repeated precisely. The planet contains more than 7 billion human inhabitants, yet no two people are precisely identical in all respects—from their mundane external features (e.g., fingerprints) to their more complex internal features (how and what they think and feel). Even the inanimate rock lying on the ground is unique. In theory, no two rocks are identical in terms of all their distinguishing characteristics.

Reality appears mostly obscure. Probably the major share of reality remains hidden from direct detection by any of our senses. To be sure, scientific instruments are being developed that enable us to probe more deeply into space, see ever tinier particles of matter and, through functional magnetic resonance imagings (fMRI), observe how regions of our brains are activated as we think, but the vast majority of nature's secrets still remains mysteriously hidden from direct view. These secrets cannot be seen, heard, tasted, smelled, or touched. With specific respect to human phenomena, whereas many are openly visible (e.g., we can see a person walking, eating a sandwich, purchasing a newspaper), a vast number of others are not. A person's psyche—the inner thoughts and feelings that presumably guide much of our behavior—is one of the most obscure realms of all. We have yet to be able to see what we think is a motive or to point to the resting place of jealousy or pride.

The four characteristics just described—that reality is experienced as complex, dynamic, unique, and mostly obscured—refer to what is often termed the *external environment*. More than a century ago, the famed philosopher and social scientist William James (1890) referred to this external environment as "a bloomin', buzzin' world of confusion." These four characteristics apply to individuals' "internal environments" as well.

CONCEPTS: THE BUILDING BLOCKS OF UNDERSTANDING

The Nature of Concepts

Confronted by this array of complex, dynamic, unique, and mostly obscured phenomena, how do individuals manage to make sense out of this world? They do so, almost automatically and usually unconsciously, by conceptualizing—that is, by using their mental processes to consider and sort their experiences in terms of the concepts they have acquired and stored in memory. They also develop new concepts to describe things they had never previously experienced. Just as concepts are the fundamental building blocks of everyday thinking, they also are the fundamental building blocks of scientific thinking.

According to *Webster's Dictionary*, the word *concept* refers to something that is conceived of in the mind. It is a generic idea or thought, usually developed from experiencing one or more particular instances. Examples of concepts include *shirt*, *book*, *dripping*, *chair*, *mother*, *ice cream*, *advertising*, *smashed*, *home*, *vacation*, *memory*, *love*, *prejudice*, *attitude*, and *expectations*. As you can see, concepts refer to things that are tangible and denotable (e.g., shirts) as well as to things that are not as concrete and directly seen (e.g., memory).

Concepts are the building blocks for all thinking, regardless of whether that thinking occurs in the context of everyday living, art, politics, sports, religion, or science. In fact, without concepts, thought as we know it would be impossible. It is our concepts that enable us to achieve some basic understanding of the world.

The most basic level of understanding can be termed *identification*. We understand something, in part, when we can identify it. When experiencing the world about us, we use the concepts we have in mind to identify and classify our experiences: This is an ice cream cone; that is a shirt. Social scientists identify and classify people using concepts such as race, gender, intelligence, and attitudes. Because concepts are so central to all thinking, we examine their nature in somewhat greater detail.

Concepts are generalized abstractions. When an individual has a concept, it means that he or she has a general idea that can be applied across a number of specific instances. Consider the concept *shirt*, for example. Shirts differ in a great number of ways—in terms of their fabric, color or number of colors, sleeve length, the number of buttons (or whether they have buttons at all), the size and shape of the collar, whether there are pockets and the number of pockets, whether the shirt is squared off at the bottom or has tails, and so on. Yet having the concept *shirt* in mind is sufficient to enable the individual to sort things into two (or possibly three) categories: shirts, items that have some of the characteristics of shirts but are not shirts, and everything else. When we say that concepts are generalized abstractions, we mean that the general idea subsumes a universe of possible instances. Note, also, that concepts can be "fuzzy" at the margins. Does a woman's blouse qualify as a shirt? What about a woman's halter top? Such fuzziness can lead to disagreements among individuals and scientists alike. For example, a recent controversy in astrophysics involved how to define the concept of a *planet*.

Concepts encompass universes of possibilities. An important feature of concepts—one that has fundamental implications for scientific theory and research—is that each concept consists of a universe of content. As just discussed, the concept *shirt* encompasses a universe of many specific possibilities. The concept *ice cream cone* encompasses a universe of possibilities. The concepts attitude toward abortion, romantic love, and so on, all encompass universes of possibilities.

Concepts are hypothetical. Concepts are not reality, just ideas regarding reality. This point is easy to appreciate when concepts or constructs apply to nebulous, amorphous, abstract things, such as *wanderlust*, *attitude*, or *sustainable development*. But this point also applies to items that are denotable and concrete. For example, although the concept of a shirt exists in our minds, we do not walk around with little shirts in our minds. Neither the word *shirt* nor the thought that this word evokes in the person's mind is a shirt. Until and unless neurological science tells us differently, concepts possess no tangible reality, in and of themselves. In this sense, all concepts are necessarily hypothetical, and, although concepts are themselves hypothetical, the things to which they refer include both observable entities (e.g., shirts, tables, dogs), which form part of the external environment, and nontangible phenomena such as love, happiness, and hunger. Although we cannot see a person's hunger directly, we can see the effects of this assumed state and, from these effects, infer its existence. Many of the concepts that populate our minds are of this indirectly observable variety.

(Most) concepts are learned. Most concepts are acquired creations. The infant does not come into the world already possessing the concept *shirt*. Rather, he or she must acquire this concept before being able to use it to understand reality and communicate with others. When individuals experience something completely new and different, they must either acquire or create a concept to be able to identify this experience and distinguish it from all other aspects they perceive. Similarly, the scientist who observes something different under the microscope or in intergalactic space will need to first conceptualize it and then give it a unique label (e.g., *chromosome*, *quasar*) with which to identify this particular phenomenon and others like it. Although most concepts are learned, there is evidence that certain concepts may be "hardwired," such as the face of a mother as perceived by a newborn (Bednar & Miikkulainen, 2003).

Concepts are socially shared. In order for communication to occur, the set of concepts possessed by one individual generally needs to be similar to the sets possessed by others. Consider trying to discuss the notions of *balks*, *punts*, and *love–15* with someone who does not understand baseball, football, or tennis, respectively. Or consider a researcher trying to discuss factor analysis with a nonresearcher who has never heard of the subject. Until both parties utilize shared concepts, communication cannot take place. That said, it is important to note that concepts in the social and behavioral sciences often have contested meanings. As examples, after reviewing the scholarly literature, Fishbein and Ajzen (1975) found more than 500 definitions of *attitude*, and Jacoby and Chestnut (1978) found more than 50 definitions of *brand loyalty*.

Concepts are reality oriented (or functional). Although not physical reality themselves, most of our concepts presumably are tied to the external world and used as a guide for interpreting and reacting to this world. Concepts are thus *functional*. If a person's interpretation and labeling of experiences do not mirror the world, then his or her reactions could be dysfunctional, even fatal. Consider the implications of conceptualizing a lethal cobra as a nonlethal garter snake. We develop and share concepts because they seem useful for helping us understand the reality we experience.

Concepts are selective constructions. The world we experience can be conceptualized in almost countless ways. For example, looking at a woman's white blouse, we can think of it as something that provides a socially expected degree of modesty, as something that offers protection from the wind and sun, as something decorative, as something that can be used to wash a car, as a bandage, as a tourniquet, or as a white flag to indicate surrender. The ways in which concepts are applied to describe reality depend upon the needs and objectives of the individual doing the conceptualizing.

Concepts, Constructs, and Variables

As might be imagined, the adult individual's mind contains a large number of concepts. Fortunately, most concepts cluster together under broader, more encompassing concepts. For example, *shirts* and *ties* are both examples of *clothing*. *Cats* and *dogs* are both examples of *mammals*, which, along with *snakes* and *insects*, are examples of *animals*. Such higher order concepts are called *constructs* because they refer to instances that are constructed from concepts at lower levels of abstraction.

We form and use constructs because they are a powerful means by which we are able to handle greater portions of reality. For example, it is much easier to say "All animals must eat in order to stay alive" than it is to say "All apes, dogs, cats, frogs, snakes, etc., must eat in order to stay alive." Not only do we use constructs because of their greater economy, efficiency and power, but also because they enable us to achieve a certain degree of order when dealing with the almost infinite number of separate concepts that populates our minds.

One type of construct that is used in many scientific theories is called a *variable*. A variable has, or is composed of, different "levels" or "values." For example, *gender* can be conceptualized as a variable that has two levels or values. That is, it can be seen as being composed of two concepts or categories, male and female. *Religion* is a variable consisting of the conceptually distinct categories Protestant, Catholic, Jewish, Muslim, and so on. *Time* is a variable that can be conceptualized as consisting of the categories 1 minute, 2 minutes, 3 minutes, and so on. *Intelligence* is a variable that has different levels ranging from low to high. Many theories in the social sciences focus on variables and the relationships between them, though the way theorists do so often differs considerably. This will become apparent in later chapters.

Variables are important because people and social entities (e.g., families, groups, organizations, nations) are thought to differ depending on the variable category or level that describes them. Males are thought to be different from females. A person with a low IQ score is thought to be different from a person with a high IQ score. Democracies are thought to be different from monarchies.

Although variables are central to many scientific theories, some theoretical approaches eschew variable-oriented approaches to theory construction. These theories tend to rely more on process-oriented characterizations of phenomena and/or on narratives (Mohr, 1982). For example, rather than thinking of gender as a variable that has two levels, male and female, these theoretical frameworks emphasize the many ways in which gender is understood by different individuals, which may include concepts such

BOX 2.1. Concepts, Cultures, and Values

Concepts are an integral part of science. They form the foundation for the way in which a scientist thinks about a problem. A major tool used by humans in categorizing phenomena is language. Several linguists, such as Benjamin Whorf, have suggested that language is more than a convenient tool for communication. Rather, language shapes the way in which people think about things. Analysis of divergent cultures clearly demonstrates that languages categorize our environment in different ways. Navajo Indians, for example, have color terms that roughly correspond to our *white*, *red*, and *yellow*, but none that is exactly equivalent to our *brown*, *gray*, *black*, *blue*, and *green*. Our *gray* and *brown* are denoted by a single term, as are *blue* and *green*. The Navajo have two different terms to refer to *black*, one focused on objects and the other on darkness. In short, the Navajo language approaches the color spectrum differently from traditional English. Differences in language do not necessarily limit the ability of an individual in one culture to "think" less well than an individual in another culture. Instead, language seems to direct perception and thought into culturally determined channels or categories. This being the case, it is evident that science is influenced by culture, including the way in which we are raised and the way we learn to categorize and relate different concepts. Many concepts are nearly universal, whereas other concepts are culture specific. In this sense, as well as others, science and scientific thought are influenced by the culture and environment in which its practitioners find themselves.

as *bisexual*, *transgendered*, and *questioning* that are not treated as levels of a variable in a broader theory. (We discuss process-oriented approaches in greater depth in Chapter 10.)

CONCEPTUAL SYSTEMS: THE BASES FOR DEEPER UNDERSTANDING

By enabling us to identify, describe, differentiate, classify, and segregate our experiences, concepts assist us in achieving some rudimentary understanding of reality. Yet, used in isolation, concepts and variables typically provide a limited degree of understanding. It is only when concepts are placed into relationship with each other that they move us toward achieving a deeper understanding of our reality. Consider the concepts *convict*, *chair*, *smashed*, and *hungry*. Although we understand what each of these concepts means separate and apart from each other, as reflected by the statement "The hungry convict smashed a chair," connecting the concepts with each other in this manner leads to seeing a number of relationships, including (1) chairs can be smashed, (2) convicts can smash chairs, (3) convicts can be hungry, and (4) hunger may cause convicts to smash chairs.

Relationships can occur on a myriad of levels. Examples of relationships include spatial relationships (e.g., the car is *on* the street, parked *next to* the sidewalk), temporal relationships (the blue car reached the traffic light *before* the green car did; adolescence *precedes* adulthood), deterministic relationships (the slip on the ice caused her to fall), kinship relationships (Jon is Robin's brother), and legal relationships (Jon is married to Beth). When two or more concepts are linked together to represent relationships, we have a rudimentary *conceptual system*. It is these conceptual systems that enable us to arrive at deeper levels of understanding.

Over the course of a lifetime, each individual tends to acquire tens of thousands, perhaps even millions, of concepts. When the number of permutations and combinations is considered, it quickly becomes apparent that each person's mind can contain a dizzying array of conceptual systems. The potential exists for all of these systems to get in the way of each other and impede understanding. This is where *selection mechanisms* come into play.

When the mental system is working effectively, concepts useful for understanding and coping with the experienced reality of the moment come into play. By analogy, it is like the situation involving the college freshman who, while standing on the steps of the administration building during his or her first day on campus, asks a passerby, "How do I get to the math building?" In answering the question, the passerby might draw a simple map to represent how he or she thinks the freshman should proceed. Clearly, the map would not depict every tree and blade of grass, every section of pavement, every building and parking lot. The only things that the map maker includes are those he or she believes will help the freshman get to the desired destination—that is, those items the mapmaker considers to be useful guides to reality. Had the freshman inquired instead "Where is the large oak tree on campus that everyone seems to be talking about?" the map might have contained different elements, even though both maps would be referring to the same physical space.

When coping with the ongoing world, our conceptual systems are analogous to a mental "map" in the previous example, though in real life the systems can take many forms, such as mental narratives, numerical representations, pictorial representations, and so on. The nature of the conceptual system that is invoked depends upon the needs of the individual at that moment. It is useful to keep this analogy in mind when we later discuss the topic of scientific theory. At that point, the reader will be able to recognize that, like maps, scientific theories are essentially conceptual systems designed to be useful in identifying, organizing, and, as discussed below, explaining or predicting some delimited portion of the experienced world.

A core facet of a conceptual system derived to provide insights into a phenomenon is what scientists call *explanation*. Although we may understand that two or more things are related, we may still not understand why this is so. Answering "Why?" involves moving to deeper levels of understanding, with the answers to such questions representing explanation. For example, why do some married couples who have been together for 20 years or more divorce? Why do some schools rely on standardized testing as an indicator of how well they are teaching their students? The answers to such questions are a form of explanation.

Another facet of understanding is being able to predict when something will happen in the future. Although prediction and explanation often go hand in hand, the two are distinct. The person who tells the auto mechanic "Every time I step on the car's accelerator, I hear a rattle in the engine—let me step on it and you can hear what I mean" may be able to predict without being able to explain. Similarly, as a moment's reflection about weather forecasting will reveal, being able to explain how the weather we experienced today came to be does not necessarily mean that we also are able to accurately predict the exact date and time when this precise weather will occur again.

Yet another feature of understanding is that it allows us to differentiate between concepts and events. As Runkel and McGrath (1972) have emphasized, knowledge is "knowledge of differences," namely, how things are similar and how things are different. Understanding a phenomenon implies we can describe what differentiates it from other phenomena or we can differentiate instances of it. Knowledge of males and females is knowledge of how males and females differ (or are similar) on different properties, dimensions, or behavior.

Thus, as used here, the term *understanding* encompasses identifying, describing, organizing, differentiating, predicting, and explaining. These basic ingredients of understanding are just as characteristic of the person on the street as they are of the scientist plying his or her profession.

Armed with an understanding of our world, we can begin to achieve important goals. These goals can be numerous and diverse, but two are especially noteworthy: *satisfaction* and *control*. Once we are able to identify, organize, and explain our experiences, the world becomes less of a frightening, unfathomable experience. Thus, if it does nothing else, understanding enables the individual to achieve a measure of peace and satisfaction. Understanding also gives us some ability to control events or relationships. Controlling the environment involves two components: (1) understanding the relevant features of the environment, and (2) having the ability to manipulate those features.

COMMUNICATION

Having traversed the terrain from concepts to conceptual systems, we are now ready for a major extension. Up to this point we have focused on what presumably happens in the minds of each of us as we try to come to grips with the world around us. But what happens when we try to communicate this understanding to another person? How do the thoughts in the mind of one person come to be represented in the mind of another? This is a particularly interesting question when we realize that human communication need not involve face-to-face verbal interaction between two people, or even communication occurring at the same time. We are still reading and benefiting from the works of Plato, and much of our communication in this Internet era occurs through written words.

When a person (whom we term the *source*) deliberately engages in communication, he or she does so because there is some thought or feeling that he or she wishes the other party (called the *receiver*) to understand. Communication is typically defined as a process whereby a source transmits a message over a medium to one or more receivers. Unfortunately, the thought that exists in the mind of a source cannot be directly transposed into the mind of a receiver. For the source to communicate a thought (i.e., evoke the intended meaning in the mind of the receiver), the source must convert it into some externally denotable form, such as the spoken word, written words, or some other detectable symbols, and convey these *symbols* to the receiver. In turn, the receiver must then decode—that is, interpret—this overt expression and extract meaning from it—hopefully, the same meaning intended by the source.

The distinction between concepts (as internal mental representations) and symbols (the external observable expressions that are used to represent internal concepts) is important. Using the vocabulary that has evolved, *meaning structure* is the term used to designate the concepts or thoughts that exist in the minds of individuals, and *surface structure* is the term used to designate the symbols that are the externally visible expression of these thoughts. In communicating a particular thought (which we label *Meaning Structure 1*), the source uses some surface structure in an attempt to evoke the same thought (Meaning Structure 1) in the mind of the receiver. Should this surface structure evoke some other meaning in the mind of the receiver (say, Meaning Structure 2 or 3 or 4), then we have a miscommunication in which meanings are not common or shared. Should the source succeed in evoking Meaning Structure 1 (MS1) but also evoke one or more other meaning structures, then we have ambiguous, confusing communication (i.e., a combination of accurate and inaccurate communication).

Suppose the source, trying to communicate that a particular automobile had surreptitiously been taken from its rightful owner, said, “This car is hot.” Although the receiver might extract the intended meaning (i.e., MS1—the car is stolen), he or she might also extract some other meaning, such as MS2—the car has just been running hard, and the engine temperature is relatively high; or MS3—the car has been sitting out in the sun and its interior, particularly the seats, would not be a comfortable place to sit; or MS4—the car has excellent high-performance characteristics; and so on. The particular set of surface structure symbols used by the source in this instance does not seem to have accurately conveyed the meaning he or she had in mind.

Just as clearly, the source could have employed any number of different surface structures to convey the intended meaning. He or she could have said, “This car is stolen,” or “This is a stolen vehicle.” Instead of speaking these words aloud, he or she could have written them out, used Morse code or the gestures employed by American Sign Language, or even tried smoke signals, pantomime, or Braille. Earlier we noted that each concept (e.g., shirt) actually represents a universe of possible meanings. Now we see that each of these meanings can be expressed via a universe of possible symbols. Since symbols tend to possess more than one meaning (i.e., they can be ambiguous), communication—including scientific communication—is enhanced by careful attention to the selection and use of symbols.

Several important points can now be summarized with respect to communication. First, it is necessary to make a distinction between people's understanding of their environment (as represented by the concepts they have in mind) and their description of that environment (as represented by the symbols, usually words, they use to describe their thoughts). Second, a number of different surface symbols could be used to communicate the same underlying meaning structure. Third, communication also requires that the receiver possess a concept comparable to the one in the mind of the source; otherwise, the communication of meaning is difficult. For example, the source could just as well have said, "The frammis is hot." If the receiver had no idea of what a *frammis* is, there would be no transference of meaning. Fourth, for any number of reasons (including "noise" in the channel), though the surface structure may have been precise and accurate, the recipient extracted an incorrect meaning. For example, the receiver may hear only part of what was said and, as a result, believe that the car is not a stolen vehicle. Fifth, even if he or she does extract the meaning intended by the source, the receiver may consider it to be inaccurate or incorrect in its description of reality (e.g., "I don't care what you tell me, that's not a stolen vehicle"). Sixth, just as the concept in our mind (e.g., of a car) is not equivalent to the elements of reality that are so conceptualized, so is the external symbol (e.g., the word *car*) not one and the same with the physical reality it describes. The symbols we use are arbitrary constructions. Finally, meaning must be interpreted in context, and a particular symbol may have different meanings in different contexts. For example, *thin* means something different when it is applied to people than when it is applied to liquids.

Of course, human communication is far more complex and dynamic than the above characterization. At this point, it is sufficient to recognize that our discussion regarding concepts and conceptual systems refers to what is happening in the mind of an individual and that to communicate with others regarding these thoughts, the individual must convert thoughts to another system that requires the use of external symbols, usually language. Since much of the scientific enterprise involves the communication of information between individuals, it is important to understand the core elements that underlie that communication. We return to this key point in later chapters when we discuss developing conceptual definitions in theory construction.

SUMMARY AND CONCLUDING COMMENTS

The world that we experience is multifaceted, dynamic, unique, and mostly hidden from direct view. At a most basic level, individuals cope with this complexity by forming and using concepts to assign meaning to (i.e., to identify and describe) their experiences. People place concepts into relationships with other concepts and use these conceptual systems as guides to organizing and explaining the world they experience. Scientists disagree about the best way to conceptualize reality, as reflected in the philosophical orientations of realism, social constructionism, critical realism, and hypothetical realism. In order to share and interact with others regarding these conceptualizations, people (scientists) translate their internal concepts into external symbols or language. When both the symbols and the underlying conceptualizations to which they refer are reasonably well shared, the exchange of meaning from one individual to another can take place.

Having described how the individual goes about achieving some measure of understanding of the world, we have provided the foundation for understanding what science is and where it fits into the world at large. In a nutshell, science is just one of a number of approaches (e.g., the arts, religion) for acquiring a deeper understanding of the world we experience. In the next chapter we examine scientific thought more formally, contrasting it with other ways of knowing things and with some of the core ideas discussed in the present chapter.

SUGGESTED READINGS

Blumer, H. (1969). *Symbolic interactionism: Perspective and method.*—Contains a cogent discussion of hypothetical realism.

Chomsky, N. (1972). *Language and mind.* New York: Harcourt Brace Jovanovich.—An in-depth discussion of surface and meaning structures in language as well as how language influences thinking.

Gergen, K. J. (1985). The social constructionist movement in modern psychology. *American Psychologist, 40,* 266–275.—A brief introduction to social constructionism.

Gergen, M., & Gergen, K. (2003). *Social construction: A reader.* London: Sage.—A collection of classic and contemporary articles that encompasses the major viewpoints of social constructionism.

Godfrey-Smith, P. (2003). *Theory and reality: An introduction to the philosophy of science.* Chicago: University of Chicago Press.—An introductory treatment to major approaches to the philosophy of science, including logical positivism and the perspectives of Popper, Kuhn, Lakatos, Laudan, and Feyerabend, as well as critiques from the sociology of science and feminism. The book also reviews scientific realism, scientific explanation, and Bayesian explanation.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by.* Chicago: University of Chicago Press.—A discussion of concepts and constructs and how we use them in our everyday life, from the perspective of anthropology.

Okasha, S. (2002). *Philosophy of science: A very short introduction.* New York: Oxford University Press.—A brief but balanced introduction to major themes in the philosophy of science.

Rosch, E. (1977). Human categorization. In N. Warren (Ed.), *Advances in cross-cultural psychology* (pp. 1–72). New York: Academic Press.—A discussion of categorization systems from a psychological perspective.

Salmon, W. (2006). *Four decades of scientific explanation.* Pittsburgh: University of Pittsburgh Press.—Traces the major debates about scientific explanation in the philosophy of science from 1845 to 1990.

Scarr, S. (1985). Constructing psychology: Making facts and fables for our times. *American Psychologist, 40,* 499–512.—Another interesting perspective on social constructionism.

Varela, F., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience.* Cambridge, MA: MIT Press.—This book is a creative account of how laypeople think about concepts and categories in everyday life, integrating research in cognitive science with Buddhist psychology.

KEY TERMS[2]

realism (p. 7)
social constructionism (p. 7)
critical realism (p. 8)
hypothetical realism (p. 9)
pragmatism (p. 9)
concept (p. 11)
constructs (p. 13)
variables (p. 13)
conceptual system (p. 14)
explanation (p. 15)
understanding (p. 16)
meaning structure (p. 17)
surface structure (p. 17)

EXERCISES

Exercises to Reinforce Concepts

1. Explain how everyday thinking is similar to scientific thinking.
2. Describe the nature of *reality* as people typically experience it.
3. What is the difference between realism and social constructionism?
4. What are critical realism, hypothetical realism, and pragmatism?
5. Define the terms *concept*, *construct*, and *conceptual system*. Explain how each contributes to understanding our environment.
6. What are the major characteristics of concepts?
7. Explain how conceptual systems function as *selection mechanisms*. How does this aid understanding through explanation and organization?

[2]For all key terms, the number in parentheses indicates the page on which the term first appears.

8. Explain how understanding the environment results in prediction and control.
9. Define what is meant by meaning structure and surface structure. Explain how they can result in different forms of communication.

Exercise to Apply Concepts

1. The United States has one of the highest teen pregnancy rates among all developed countries in the world. A social scientist wants to better understand teenage pregnancy in the United States. How would the material in this chapter shape the way in which the scientist might think about this topic?

3

Science as an Approach to Understanding

The work of science is to substitute facts for appearances, and demonstrations for impressions.

—JOHN RUSKIN (1859)

Chapter 2 described how individuals use concepts and conceptual systems to achieve an understanding of the world, with the premise that scientists draw upon many of these same strategies to construct scientific theories. In this chapter we delve more deeply into scientific thinking and theorizing. We begin by considering different socially based approaches to understanding, such as theology, philosophy, jurisprudence, the arts, literature, and science. We describe the key characteristics that separate science from these other "ways of knowing." Next, we discuss key concepts in theory construction, including the definition of a theory; the difference between theories, models, and hypotheses; and the different typologies that scientists use to characterize theories. We conclude with a discussion of the characteristics of a good theory and the role of objectivity in science.

SOCIALLY BASED APPROACHES TO UNDERSTANDING

There are many ways of gaining and organizing knowledge about one's world, only one of which is science. These socially based approaches to understanding all involve internal conceptual systems that are communicated among individuals using an externally observable shared symbol system (e.g., words, gestures, mathematics). Being able to use such shared symbol systems opens the door to opportunities for improving and expanding personal understanding. Not only does it enable the individual to communicate his or her thoughts *to* others, but also to receive communications *from* others. These others may have a more useful way of looking at the world that may lead the individual to revise his or her thinking. In addition, communicating thoughts to others can help individuals clarify their logic through self-reflection during the communication process.

The existence of shared symbol systems also enables us to tap into the accumulated wisdom of the past. After all, a great number of the things that each of us experiences has been experienced, thought about, and discussed by others at earlier points in time. It is possible that we would find these conceptualizations useful in our own attempt to understand our world. We refer to these bodies of knowledge as *shared meaning systems*. For the present, this term is meant to refer to both the underlying conceptualization as represented in our minds and the externally visible symbols used to communicate regarding this conceptualization.

There are many examples of shared meaning systems. Mythology documents myths that are or were used to explain otherwise unexplained natural phenomena (e.g., the sun dropping out of view in the evening and mysteriously reappearing the next day), imbuing these phenomena with meanings that made them appear less mysterious. Other perspectives also have evolved over the ages, including those in the diverse fields of theology, philosophy, jurisprudence, the arts, literature, and science, to name a few. Each reflects a different orientation to ordering and understanding the world we experience. The fact that these perspectives have persisted for centuries suggests that each provides a satisfying way of extracting meaning from, and coping with, the world for significant numbers of people. In some key respects, science is like any of the other approaches mentioned above.

Commonalities across All Shared Conceptual Approaches

At least three fundamental characteristics appear to characterize all shared approaches to understanding. First, each approach consists of concepts and relationships among these concepts. In this regard, all shared approaches—including science—are like the conceptual systems used by individuals. The basic difference is that the shared systems tend to be more elaborate, more abstract, stabler over time, and more explicit.

Second, all shared belief systems are limited in how much of the world they address. Indeed, if they possessed no such limitation, they would be forced to grasp all of the complexity of the ongoing world as it progressed, and that would be impossible. As a consequence, no single orientation (including science) has an exclusive franchise on arriving at exhaustive and comprehensive understanding. The Nobel Prize-winning physicist Weisskopf (1977, p. 411) observed:

> Human experience encompasses much more than any given system of thought can express. . . . There are many ways of thinking and feeling, each of them contains some parcel of what we may consider the truth. . . . Science and technology comprise some of the most powerful tools for deeper insight and for solving the problems we face . . . but science and technology are only one of the avenues toward reality: others are equally needed to comprehend the full significance of our existence.

Recognition of the limits of science has been expressed in many ways. Consider what the psychologist Scarr (1985, p. 500) said:

> Science, construed as procedures of knowing and persuading others, is only one form of knowing by the rules of one game. There are other games in town, some like art more intuitive, some like religion more determined by revelation and faith.

A third feature of shared belief systems is that they generally serve prescriptive and evaluative functions. The *prescriptive function* can be thought of as guidance regarding how we *ought to* approach or respond to some aspect of reality or experience. The formal systems of religion, for example, may provide explicit guidance on such subjects as premarital sex and birth control. In certain instances, the formal system of science indicates what are and are not proper procedures to be followed. For example, scientists should subject theoretical propositions to empirical tests to gain perspectives on the viability of the propositions. Prescription provides a basis for evaluation. Given that we have some notion of what should be done, we can evaluate how well what has been done corresponds to what should have been done. The *evaluative function* permits labeling something as being good or bad, right or wrong. Shared systems thus generally provide a template or model against which to evaluate activities that purport to have been taken in accordance with that model.

Special Features of the Scientific Approach

If all shared belief systems consist of the same underlying foundation (concepts and relationships) and each can accommodate only limited portions of our environment, then what distinguishes science from these other approaches? The answer has to do with how the worth of the statements and inferences within the system is assessed. To be taken seriously, any shared system needs to demonstrate that it possesses utility, that is, that it provides some useful way of describing or coping with the world about us. A variety of avenues is available for assessing the usefulness of a conceptualization. Perhaps the most common strategy is that known as *consensual validation*. In this approach the worth of a particular conceptualization is gauged by the degree of acceptance it is granted by others. The fact that other people believe that a particular conceptualization is correct is used as the basis for contending that it necessarily is correct. For example, in the legal context, if a jury agrees that some particular view must be correct (and this view has been sustained on appeal), then its verdict is accepted as being correct within the legal system. Consensual validation also typifies those religions where gaining and retaining adherents are interpreted as bearing on the validity of the underlying tenets. The fact that many believe in the religion is interpreted as an indication of its validity, since “so many people could not be wrong.” Consensual validation also surfaces in the arts, where public acceptance serves to validate an artistic conceptualization.

Expert validation is a related avenue for assessing the value of a particular conceptualization. Here, the decision as to whether a particular conceptualization merits

BOX 3.1. The Fringes of Science

Scientific theories are subject to many types of validation. Several critics of science believe that scientists are overly zealous in their application of one type of validation, consensual, and that this can hamper the advancement of knowledge. These critics contend that scientists are too quick to dismiss researchers and theorists "working on the fringes" and that the scientific community only takes seriously that which is acceptable to the prevailing views of that community. Stated another way, science is inherently conservative. As a result of relying on such consensual validation, there are many missed opportunities. A frequently cited example is that of Galileo. Based on his observations with the recently invented telescope, Galileo came to question many widely held beliefs about the universe, such as the Earth being at its center. Ultimately, much of what Galileo posited proved to be true, even though he was subjected to public ridicule and brought before the Inquisition in Rome.

It is, of course, true that strict adherence to prevailing views may blind the scientist to new insights and advances. But this does not mean that the scientific community should approach "fringe" claims without a healthy skepticism. Consider, in retrospect, the case of Galileo. Even though Galileo was ridiculed by the general public, his observations were carefully scrutinized by the scientific community. With the invention of the telescope, scientists had no way of knowing whether what could be seen through the lens was, indeed, accurate. At the time, there was no theory of optics, and what is taken for granted today about the behavior of glass lenses was unknown at that time. Galileo asserted the validity of his telescope by examining objects on Earth with it and demonstrating its accuracy when compared to the case where the objects were directly observable to the human eye. Unfortunately, some distortions occurred at times, such as double images and color fringes. In addition, Galileo observed that while the telescope magnified planets and moons, fixed stars appeared *smaller* in size. Without a theory of optics, Galileo was unable to explain these phenomena, and, as such, the scientific resistance to his ideas may not have been as irrational as commonly portrayed.

Scientists must think carefully about the factors that influence their judgments regarding the validity of a theory, and be explicit about the criteria that they are using when evaluating that theory. Many of the "fringe theories" that occur in the popular press (e.g., biorhythms, the Bermuda triangle) simply do not hold up under careful empirical evaluation, despite claims by their adherents of being treated like Galileo.

acceptance is determined by selected others who presumably have the knowledge and wisdom to discern what is and is not correct. Examples include relying on professional critics to determine the validity of artistic conceptualizations, on judges to decide the truth of legal matters, and on religious leaders to decide the truth of religious conceptualizations. Another confirmation strategy, *internal validation*, involves the application of formal rules of logic to examine the concepts and relationships within a particular conceptual system. If these concepts and relationships withstand the rigors of intensive logical assessment, then the conceptualization is said to be confirmed. Such a confirmation strategy is typically employed in philosophy and mathematics.

Although science also employs consensual, expert, and internal validation, the scientific approach can be differentiated from all others by the fact that it is the only one to place primary reliance on *systematic empirical validation*. Over the long run, scientific conceptualizations tend to be accepted only to the extent that they have been subjected to rigorous and systematic empirical testing and shown to be useful. We now consider this important point in greater depth.

THE ESSENTIALS OF SCIENTIFIC ENDEAVOR

At its core science can be thought of as consisting of a *conceptual realm*, on the one hand, and an *empirical realm*, on the other. The conceptual realm entails the development of a conceptual system (consisting of concepts, constructs, and their relationships) that can be communicated unambiguously to others. The empirical realm refers to the process whereby the worth of the conceptualization is assessed through the conduct of scientific studies. For example, an organizational scientist might suggest a theoretical proposition (in the conceptual realm) that female applicants who are pregnant will be less likely to be hired for jobs than female applicants who are not pregnant. The scientist then subjects this proposition to an empirical test (in the empirical realm) by designing a study to discern if such bias actually occurs. For example, managers might be asked to evaluate videotapes of applicants with identical credentials and identical interview behavior, with the only difference between them being that one applicant is obviously pregnant and the other is not (for such a study, see Cunningham & Macan, 2007, who found evidence for such a bias).

Regardless of how detailed, formally explicit, or elegant they may be, by themselves, conceptual systems (such as theories, models, and hypotheses) are not scientific, only prescientific. To be fully scientific, they need to be subjected to empirical testing. As Carnap (1936, 1937), Pap (1962), and Popper (1963) have argued, "a scientific statement that claims to say something about the actual world . . . is meaningful if and only if there are possible observations whose outcome is relevant to the truth or falsehood of the statement" (Pap, 1962, p. 24). Science avoids *metaphysical explanations*, that is, conceptualizations that cannot be publicly observed and tested. For a conceptual system to be considered scientific, corresponding efforts must be generated toward its empirical verification (or falsification; see Popper, 1968).

As an example, the concept of unconscious influences on behavior had a major impact on psychological theory when the unconscious was first popularized by the theories of Sigmund Freud. However, scientists soon became skeptical of the use of constructs about the unconscious because the constructs could not be validly measured and statements about them could not be subjected to empirical verification. It was possible for constructs about the unconscious to be invoked post hoc to explain most any behavior; without the possibility of empirical tests, such explanations could never be falsified. Interestingly, there has been a resurgence in the study of constructs about the unconscious as predictors of behavior as new technologies have become available that purportedly allow for the measurement of facets of the unconscious (e.g., Blanton & Jaccard, 2008).

Just as theoretical propositions must be addressed in the empirical realm for science to progress, it also is the case that empirical systems make little sense without a corresponding conceptual system to organize them (see Kaplan, 1964, pp. 159–161). Any phenomenon or environment can be thought of as consisting of a great number of empirical "facts." "Without some guiding idea, we do not know what facts to gather" (Cohen, 1956, p. 148). It is often said that research should be pursued without preconceived ideas. This is impossible. Not only would it make every research investigation pointless, but even if we wished to do so, it could not be done (Poincaré, 1952, p. 143). Even when collecting "facts," we must have some hypothesis or guiding ideas as to which facts are relevant to the investigation at hand, since we can hardly amass all the facts in the universe. A researcher interested in understanding the bases of poverty in the Maya living in the highlands of Guatemala cannot randomly collect information about the Maya to address this matter. Rather, the investigator thinks about different ways of gaining perspectives on the issue and, in doing so, inescapably imposes a conceptual system, no matter how rudimentary it might be, onto the problem at hand. The basic point is that no observation is purely empirical—that is, free of any ideational element (Kaplan, 1964, p. 48). We return to this point in greater detail in Chapter 10 on grounded theory.

The necessity for both conceptual and empirical systems cannot be overemphasized. Scientific theories do not constitute science until and unless they are (or at least can be) subjected to empirical testing. Without such tests, scientific theories represent only precursors of science, that is, propositions that have not been subjected to empirical evaluation. Correspondingly, even the most applied researcher interested only in answering the question of the moment cannot escape the fact that, regardless of how latent, some form of conceptualization precedes and guides the data he or she collects and the interpretation he or she derives. The emphasis on empirical confirmation or disconfirmation and the process by which this is accomplished are the sine qua non of science and distinguish it from all other approaches to generating understanding. Note that scientists also employ consensual, expert, and internal validation, and the nonscientific approaches may point to empirical phenomena as providing confirmation for their views. However, scientists recognize that the other forms of validation do not suffice for labeling a conceptualization as *scientific*. By the same token, though the nonscientific approaches may gather data in an attempt to confirm their conceptualizations, the rigor and controls reflected by systematic empiricism generally are absent.

THE PROCESS OF THEORY CONSTRUCTION

What Is a Theory?

As described in Chapter 2, as nonscientists we develop and use conceptual systems to better understand the physical and social world around us. When working as scientists, we do the very same thing. Such conceptualizations may be based on what we observe, imagine, or are stimulated to think about after engaging in mind games of our own, considering what others have said about the issue at hand, or after empirical observations have been made. The conceptualization is then given concrete expression via some external symbol system. That is, our ideas are converted into words, numbers, diagrams, and so on. The process of formulating conceptual systems and converting them into symbolic expressions is termed *theorization* or *theory construction*.

The term *theory* has been defined in a multitude of ways by social scientists. Some examples are:

> A theory is a symbolic construction. (Kaplan, 1964, p. 296)
>
> It will be convenient for our purposes to define a theory simply as a set of statements or sentences. (Simon & Newell, 1956, p. 67)
>
> Basically, a theory consists of one or more functional statements or propositions that treat the relationship of variables so as to account for a phenomenon or set of phenomena. (Hollander, 1967, p. 55)

Although theories differ in many respects, we contend that, at their core, all theories consist of concepts and relationships between those concepts. For this reason, it is sufficient for the purposes of this book to define a theory very simply: A *theory* is a set of statements about the relationship(s) between two or more concepts or constructs.

Theories, Models, and Hypotheses

A term often used by scientists when referring to the conceptual realm is *model*. The distinction between theories and models in the social science literature is not always apparent. As examples, various authorities contend that models are a *special type* of theory (e.g., Coombs, Dawes, & Tversky, 1970, p. 4; Kaplan, 1964, p. 263), that models are *portions* of theories (Sheth, 1967, p. 444; Torgerson, 1958, p. 4), that models are *derived from* theories (e.g., Pap, 1962, p. 355), that models are *simplified versions* of theories (e.g., Carnap, 1971, p. 54), that models represent *correspondence between* two or more theories (Brodbeck, 1968), or that theories represent *specific interpretations* of (i.e., are derived from) models (e.g., Green & Tull, 1975, p. 42). Others consider the terms to be synonymous (cf. Dubin, 1976; Simon & Newell, 1956). Although there may indeed be meaningful distinctions between theories and models, it also is the case that models, like theories, involve concepts and relationships between concepts. Accordingly, we use the terms *theory* and *model* interchangeably. A *theoretical expression* refers to any external symbolic representation of an internal conceptual system, regardless of whether that symbolic representation is more properly considered a *theory* or a *model* by others, and regardless of whether the representation is verbal, mathematical, pictorial/graphic, or physical.

Another term frequently used in scientific theorizing is *hypothesis*. The nature of a hypothesis, relative to theories and models, also is somewhat ambiguous in texts on research methods. Many scientists define hypotheses as empirically testable statements that are derived from theories and that form a basis for rejecting or not rejecting those theories, depending on the results of empirical testing. For example, a researcher might want to the test the theory that people can better recall negative information about a person than positive information. This general proposition is translated into a hypothesis or prediction about what will happen in an experiment where college students are read a list of positive and negative adjectives (prechosen to occur with equal frequency in the English language) and asked to recall the adjectives 2 minutes later. The hypothesis is that the number of negative adjectives recalled by the students will be greater, on average, than the number of positive adjectives recalled. This hypothesis, stated in a form that is part of an empirical evaluation of a theory, was derived from the more general theoretical expression that the theorist wants to evaluate. Others define a hypothesis as a theoretical statement that has yet to be empirically validated. For example, the proposition that "people can better recall negative information about a person than positive information" would be termed a *hypothesis* until it has been subjected to formal empirical testing.

Like theories and models, hypotheses are statements that involve concepts and relationships between them. For this reason, we do not distinguish them from theoretical and model-based statements. All three types of conceptual systems—theories, models, and hypotheses—can be classified for our purposes under the more generic term *theoretical expression*. Henceforth, the terms are used interchangeably in this book, with a full recognition that other social scientists may make distinctions between them.

Types of Theories

Philosophers of science have developed typologies of theories so as to better understand the range of theoretical expressions that occur in science. Examples include Albert Einstein's (1934) distinction between constructive and principle theories, Marx's (1951) distinction between reductive and constructive theories, and Kaplan's (1964) distinction between concatenated and hierarchical theories at either molar or molecular levels. More recently, theories have been characterized as being humanistic, behavioristic, constructionist, structuralist, functionalist, and so on.

Although all theories focus on concepts and relationships between concepts, theories in the social sciences differ in the fundamental assumptions they make about human behavior. These assumptions lead theorists to think about the same problem in different ways. For example, a humanist may identify and conceptualize an entirely different set of concepts when analyzing school performance in children than the concepts that a behaviorist might consider. The humanist might focus on concepts such as how the child construes the school environment, the child's feelings about school, and the affective quality of the relationship between the teacher and the student. In contrast, the behaviorist might focus on the positive and negative reinforcers that the child is receiving and the nature of contingencies between performance of behaviors and the administration of rewards and punishment. Neither conception is more "correct" than the other, although one theoretical approach ultimately might satisfy the criteria of what constitutes a good scientific theory better than the other. We view broad-based typologies of theories, such as those mentioned above, as different launching points for identifying concepts and relationships that we use to organize and understand our world. We discuss such perspectives in Chapter 11.

The Role of Theory in Basic versus Applied Research

An often-heard distinction is that between basic and applied scientific research, yet the essential difference between these two types of research is difficult to identify. According to one perspective, basic researchers use theories whereas applied researchers do not. Yet every scientist, even the "strict empiricist," cannot escape the fact that, regardless of how hidden, some form of conceptualization precedes and guides the data that he or she collects and the interpretations he or she derives from it. Hence, reliance upon theory would appear to provide an unsatisfactory basis for distinguishing applied from basic research.

Another basis for distinguishing the two approaches emphasizes the intent of the researcher. When the intent is to address and hopefully solve an immediate real-world problem, the research is considered to be *applied*. In contrast, research conducted for the purpose of extending the boundaries of our collective body of understanding, not for the purpose of addressing a pressing problem, is termed *basic*. Theories are seen as being oriented toward basic or applied phenomena, depending on research objectives. According to this view, the applied and basic researcher could design and implement virtually identical studies yet, because of different research objectives, one would be termed *applied* and the other *basic*.

Another criterion that often is suggested for distinguishing between basic and applied research focuses on the abstractness of the concepts in the conceptual network. According to this perspective, applied research is typically concerned with relatively narrow and circumscribed concepts that are domain specific. For example, the blue jeans manufacturer interested in expanding sales might commission a study to determine whether the buying public contained a sufficient number of people ready for jeans in new colors, styles, and patterns. However, though interested in learning more about such innovators, he or she most likely would not be interested in funding research to learn whether respondents were also innovators in regard to other consumer products (e.g., appliances, pens, foods). Understandably, the objective is to achieve some understanding of a concrete and limited problem. In contrast, basic research is typically interested in broader, less concrete concepts. In the present instance, basic researchers would likely strive to understand and draw inferences regarding innovators in general (i.e., across the range of consumer products) and how these innovative tendencies might be related to a broad spectrum of other concepts and constructs, usually ones that have been suggested and perhaps explored in prior research by others.

There seems to be no single basis that proves sufficient for clearly distinguishing between basic and applied research. Perhaps the best approach is to note a set of attributes that, when employed in combination, seems to provide some basis for making such a distinction. From this perspective, *applied research* can be characterized as research that focuses on an immediate problem, relies on concepts that are relatively narrow in scope, and produces results that are not intended to extend a general body of knowledge. In contrast, *basic research* is characterized as research that is not directly focused on pressing real-world problems, tends to rely on concepts that are relatively broad in scope, and produces findings with the intent of contributing to and extending our basic understanding of the phenomenon in question. Regardless of whether the research is characterized as basic or applied, however, both types of research necessarily begin with some sort of conceptual system. For more extended discussions of applied versus basic theory and research, see Brinberg and McGrath (1985) and Brinberg and Hirschman (1986).

CHARACTERISTICS OF A GOOD THEORY

How do we know if a theory is a good theory? Several criteria have been proposed for evaluating theoretical expressions. If we assume that the purpose of a theory is to help us better understand our world, then the paramount consideration is whether it does indeed offer such guidance. From this perspective, the primary evaluative criterion is *utility*. Theoretical expressions are valued to the extent that they serve as useful guides to the world we experience, that is, to the extent that they enable us to achieve some understanding of our world. Recognize that utility is a relative notion. Consider being adrift in the ocean with a leaky life raft. Unless a better life raft is available, we would be foolish to discard the one that leaks—it is the best we have. As another example, though a hand-drawn map may not be 100% accurate, it may be sufficiently accurate to be useful. If a theory is flawed in some respect but still provides unique and useful insights in other respects, it tends to be retained until something better comes along.

Consensual validation is one basis by which theories are accepted or rejected by scientists. This term refers to the degree of consensus among the scientific community about the validity of the theory. If a theory enjoys widespread acceptance, then it is seen as being a "good" theory. The philosopher Karl Popper (1968) believed that adherents of what most scientists judge to be a "bad theory" eventually die off or leave science, rendering the theory obsolete with time. Shaw and Costanzo (1982) distinguish two broad classes of criteria for determining a good theory: those criteria that are necessary if the theory is to be accepted by the scientific community, and those that are desirable but not essential to acceptance. In the former category, three criteria are crucial. First, internally, the theory must be logically consistent; that is, the theoretical statements within the conceptual system must not be contradictory, nor must the theory lead to incompatible predictions. Second, the theory must be in agreement with known data and facts. Third, the theory must be testable; that is, a theory must ultimately be subject to empirical evaluation. The previously discussed constructs about the unconscious, as introduced by Freud, provide an example of an untestable theory.

In the second category discussed by Shaw and Costanzo (1982) are six criteria. First, a theory should be stated in terms that can be understood and communicated to other scientists. Second, the theory should strive to be parsimonious in that it adequately explains a phenomenon, but with a minimum of concepts and principles. Scientists refer to this criterion as Occam's razor, named after the 14th-century English philosopher William of Occam, which "cuts away" extraneous concepts and assumptions so as to yield a theory that is parsimonious yet satisfactory in its level of explanation. All other things being equal, preference is given to theories that make fewer assumptions. The fewer the working parts necessary to get the job done, the better the theoretical system. Third, while recognizing that theories are occasionally so novel that they upset the theoretical applecart, a theory should be consistent with other accepted theories that have achieved consensus among the scientific community; that is, it should be able to be integrated into existing bodies of theory.

A fourth desideratum is scope. Other things being equal, the greater the range of the theory (i.e., the more of "reality" that it encompasses), the better it may be. Though both Newton's and Einstein's theories of gravity enable us to understand a great many of the same things, the fact that Einstein's theory enables us to understand much more makes it a more powerful and valuable theory. That said, there are times when narrow-range theories tend to hold up better over time than broader-range theories. Also, as discussed in Chapter 4, scientific progress is often achieved by narrowing the focus of theories, not broadening them. Thus, this criterion is somewhat controversial among scientists and can be viewed as a two-edged sword.

Creativity or novelty is a fifth criterion sometimes suggested for evaluating a theory. A theory that explains the obvious is generally not as highly valued by the scientific community as one that provides a novel insight into an interesting phenomenon. Finally, many scientists suggest that a good theory is one that generates research activity—which often is a consequence of consensual validation of the theory. A theory that is rich in scope, explicit, interesting, and useful will probably generate a good deal of empirical research. Hence, a yardstick of a good theory is the amount of research it generates. Note, however, that some scientists (e.g., Skinner, 1957) have questioned this criterion, noting that many a theory has led investigators into research enterprises that have been a waste of time.

Brinberg and McGrath (1985) note that various theory desiderata sometimes conflict with each other. For example, parsimonious theories tend to be more limited in scope. As such, theorists often must make tradeoffs as they construct theories to maximize what is valued by the scientific community.

SCIENCE AND OBJECTIVITY

It is often asserted that scientists are objective in their approach to understanding and that the hallmark of science is its *objectivity*. In some respects, science is anything but objective. Whether consciously or not, the scientist brings to any setting a prior schema (or set of thoughts, beliefs, and assumptions) that is used to filter, interpret, and analyze the world about him or her. This is an inevitable feature of human nature and human thinking. The scientist's schema influences the selection and formulation of problems the scientist decides to study, the types of strategies the scientist uses to collect data (since such acts ultimately are determined by how a problem is formulated), and how data are interpreted so as to alter or fortify the scientist's initial conceptualization.

If, at its very core, science has such subjective characteristics, from where does its reputation for objectivity come? The objectivity of science stems from the fact that the scientist's conceptualization has a corresponding external representation that makes that conceptualization available to others so that they can scrutinize, evaluate, and repeat (or *replicate*) the work of the originating scientist. It is not necessary that other scientists agree on what the implications of these empirically verifiable facts mean. What is critical is that other scientists agree on their empirical existence and could, if they so desired, reproduce them. This characteristic of science has been termed *intersubjectivity* (Kaplan, 1964, pp. 127–128; Babbie, 1973, pp. 18–19). The enterprise of science is predicated upon a foundation of intersubjectivity; in this sense, it is objective.

Although science is heavily influenced by the conceptual schemes the scientist brings to the table, there also are aspects of the scientific enterprise that are consistent with the spirit of objectivity. In the words of Blumer (1969), science attempts to yield perspectives on the obdurate character of our social and physical environment. In doing so, scientists subject their propositions to empirical tests to try to determine the validity and utility of their statements. They strive to do so in ways that do not bias or prejudge the outcomes of their empirical tests, though they may not always be successful in accomplishing this goal. They consider competing conceptual schemes that lead to opposite predictions and then give preference to the schemes whose predictions (and hence utility) follow from the empirical tests. Although pure objectivity is rarely achieved, it still represents a working goal for many scientists, the pursuit of which functions to help scientists choose between conceptual schemes in terms of their relative utility.

SUMMARY AND CONCLUDING COMMENTS

By themselves, individuals are limited in the amount of their environment with which they can cope and understand. Their power is increased many-fold when they incorporate the efforts of others in this regard. To acquire such deeper levels of understanding, the individual typically relies on shared conceptual systems. A number of different shared conceptual approaches exist, including religion, philosophy, law, music, art, and science. Despite their unique variations, all conceptual systems can only provide partial understanding. Each is capable of providing a unique perspective, which may reinforce or expand upon the understanding generated by the others.

Science is distinguished from all the other shared conceptual approaches by the strategy it favors for evaluating its conceptual systems. This strategy, known as systematic empirical confirmation, requires gathering (or, more accurately, generating) relevant information from external observations that are capable of being verified or disproved by observations made by others. In turn, systematic empirical confirmation is predicated upon theorizing. Theorizing involves conceptualizing some phenomena in terms of a set of expressions, encompassing concepts and relationships among them, and then expressing these ideas via a symbol system, typically words and/or numbers. Scientists have described a range of criteria for evaluating theories, some of which are deemed essential whereas others are deemed desirable. The process of theorizing is a complex enterprise that is difficult to teach. The remainder of this book provides the reader with heuristics and conceptual systems that may prove useful in such endeavors.

SUGGESTED READINGS

Ben-Ari, M. (2005). *Just a theory: Exploring the nature of science.* New York: Prometheus.—A computer scientist explores the elements that qualify something as science as well as the characteristics of a valid theory.

Brinberg, D., & McGrath, J. (1985). *Validity and the research process.* Newbury Park: Sage.—A classic on the research process that includes useful discussions of conceptual systems.

Einstein, A. (1934). *Essays in science.* New York: Philosophical Library.—A discussion of types of theories and the process of theorizing.

Haack, S. (2007). *Defending science—within reason: Between scientism and cynicism.* New York: Prometheus.—A probing account of how science interacts with, and is influenced by, other areas of human endeavor, including literature, jurisprudence, religion, and feminism.

Hempel, C. G. (1966). *Philosophy of natural science.* Englewood Cliffs, NJ: Prentice-Hall.—A classic book on the general topic of the philosophy of science.

Kaplan, A. (1964). *The conduct of inquiry.* San Francisco: Chandler.—A thoughtful analysis of the philosophy of science.

Lakatos, I., & Musgrave, A. (1970). *Criticism and the growth of knowledge.* Cambridge, UK: Cambridge University Press.—A critique of several popular conceptions of the philosophy of science.

Loker, A. (2007). *Theory construction and testing in physics and psychology.* Victoria, Canada: Trafford.—A comparison of theory construction approaches in physics with those in psychology, offering a somewhat jaundiced view of the latter.

Parsons, K. (2004). *The science wars: Debating scientific knowledge and technology.*—An edited volume covering a wide range of social and cultural influences on the practice of science.

Popper, K. (1968). *The logic of scientific discovery.* London: Hutchinson.—A classic on the philosophy of science.

KEY TERMS

shared meaning system (p. 23)
prescriptive function (p. 24)
evaluative function (p. 24)
consensual validation (p. 24)
expert validation (p. 24)
internal validation (p. 26)
conceptual realm (p. 26)
conceptual system (p. 26)
metaphysical explanation (p. 26)
theory construction (p. 28)
theory (p. 28)
model (p. 29)
hypothesis (p. 29)
basic versus applied theorizing (p. 30)
intersubjectivity (p. 33)

EXERCISES

Exercises to Reinforce Concepts

1. Explain how shared belief systems are useful in our attempt to understand the environment.
2. Identify and explain the three fundamental commonalities of shared belief systems.
3. Explain how science is similar to other belief systems. How does it differ?
4. Distinguish between prescriptive and evaluative functions.
5. Identify and explain different ways of evaluating the usefulness of conceptualizations. Which of these are emphasized by the scientific approach?
6. Identify and define the two basic realms of science. Which is more important and why?
7. What are the characteristics of a good theory?

8. "The strength of a chain is determined by its weakest link." Explain how this concept applies to scientific research.
9. Explain what is meant by the intersubjectivity of science.

Exercise to Apply Concepts

1. From the literature of your choosing, find a theory and describe it. Evaluate that theory using the major criteria discussed in this chapter for evaluating theories. If you have difficulty applying one of the criteria, describe why. Identify other criteria, if any, you might use other than those discussed in this chapter.

Part II
CORE PROCESSES

4

Creativity and the Generation of Ideas

The difficulty lies not in the new ideas, but in escaping the old ones.

—John Maynard Keynes (1936)

The Nobel Prize-winning scientist Murray Gell-Mann was asked by a prospective student at the California Institute of Technology if the school taught the problem-solving methods used by the brilliant Nobel Prize-winning physicist Richard Feynman, also a faculty member at the university. Gell-Mann replied "no," and when the student asked why not, he responded: "Here is Feynman's method. First, you write down the problem." Gell-Mann then squeezed his eyes closed and put his fists against his forehead. "Second, you think really hard." Opening his eyes, he ended by saying: "Third, you write down the answer."

Because Feynman's method probably will not work well for you, the present chapter provides more concrete guidance for theory construction. Theory construction involves specifying relationships between concepts in ways that create new insights into the phenomena we are interested in understanding. As we seek to explain something, we do so by invoking concepts and processes that we think influence it or are the basis for it. For example, to explain why some children perform poorly in school, we might try to think about the characteristics that discriminate good performers from bad performers. When making a list of these characteristics, we will, in essence, identify constructs that are related to school performance. Or, we might want to explain why so many indigent Mayan Indians living in the highlands of Western Guatemala are converting from Catholicism to Protestantism. Again, we might try to think about the characteristics that are unique to recent converts, and as we list these attributes, we will be identifying variables or constructs that are related to conversion.

A first step in theory construction often is one of generating ideas about new explanatory constructs and the relationships between them or generating ideas about the mechanisms underlying the phenomena that you are trying to explain, without initially being too critical about the merits of these ideas. The ideas generated are then subjected to more careful analytic scrutiny, with "bad ideas" being rejected and promising ideas being pursued further. As you choose the key constructs and relationships to focus upon, you need to refine them so that they meet the rigors of a formal scientific theory. Chapters 5 and 6 discuss strategies for refining and focusing concepts and relationships. The present chapter considers the initial process of idea generation.

There is no simple strategy for generating good ideas or good explanations. It is a creative process that is difficult to articulate, describe, and teach. In this chapter we briefly review research on creativity to give you perspectives on the mental and social processes involved in the creative process. Next, we describe issues to consider when choosing a topic or problem to study. We then describe concrete heuristics that may help you generate ideas. Finally, we describe creative strategies used by influential innovators in the social sciences. Chapters 7 through 11 build on the material in the present chapter in a more substantive way. The present chapter is just a start.

ONE SMALL STEP FOR SCIENCE

Name the first great scientist–theorist who comes to your mind. Perhaps it is Albert Einstein. Perhaps it is Isaac Newton. Perhaps it is Sigmund Freud. All of these individuals had a monumental impact on their respective fields of study. Theoretical advances in the social sciences do not, of course, require revolutionizing the field in the way that these individuals did. There is ample room for the more typical yet useful small increments in knowledge that solid theoretical work and research offer (Kuhn, 1962). Indeed, the gradual building of knowledge is an essential aspect of the scientific endeavor. As they attempt to explain variation in behavior, some scientists chip away at answers with the scientific tools equivalent to a small hammer and chisel. Gradually, small bits of knowledge cumulate into larger groupings of knowledge, and eventually we gain a sense of why some people behave one way and others behave another way.

On the other hand, thinking "big" should not be avoided. Instead of approaching explanation with a chisel and small hammer, some theorists prefer to use the scientific equivalent of a sledge hammer, knocking away large chunks of unexplained variation in behavior by focusing on fundamental, pervasive, and important processes. As we discuss below and in Chapter 13, there are forces operating in the scientific environment that reward and punish both approaches.

CREATIVITY

This section briefly reviews research on creativity to provide a sense of the processes that are involved when thinking creatively. After reviewing this research, we extract practical implications for constructing theories.

The Creative Person

Sternberg and Lubart (1996) define creativity as the ability to produce work that is both novel (original or unexpected) and appropriate (useful or meets task constraints). Early studies of creativity focused on the concept of genius and the lives and minds of eminent artists, writers, and scientists. For instance, research in the 1950s and 1960s attempted to identify the personality characteristics of highly creative individuals in different fields, including science, mathematics, writing, architecture, and art. As one example of this, Frank Barron of the Institute for Personality Assessment and Research at the University of California at Berkeley used nomination techniques to identify outstanding creative writers and invited them to participate in extensive testing and interviewing sessions. Some of the writers who participated were world renowned (e.g., Truman Capote, Norman Mailer, W. H. Auden). Barron found that the personalities of creative writers were characterized by independence and nonconformity, drive and resiliency, risk taking, ambition, a concern with philosophical matters, frankness, social activism, introversion, depression, empathy, intensity, a heightened sense of humor, and trust in intuition (Barron & Harrington, 1981).

Nakamura and Csikszentmihalyi (2001) conducted an intensive case study of Linus Pauling, the eminent scientist whose valence-bond theory had major implications for the science of chemistry. These authors used the life of Pauling to illustrate that creativity is not just a product of intrapsychic processes but that it fundamentally involves the incorporation of novelty into culture. Creative contributions are the interaction of three systems: (1) the innovating person, (2) the substantive domain in which the person works, and (3) the field of gatekeepers and practitioners who solicit, discourage, respond to, judge, and reward contributions. At the person level, Pauling had characteristics that increased the likelihood of creative contributions, including intense curiosity and a love of science, a quick, playful mind, and an incredible memory that enabled him to draw on vast knowledge bases. He was adept at imaging, which allowed him to analyze complex dimensional structures more efficiently than the typical person. He also had strong mathematical skills that were needed for analyses in quantum physics. Pauling liked to think about the bigger picture. Importantly, he was gifted at explaining complex ideas in clear and simple terms. It was the latter qualities that helped him persuade the field to accept his ideas. As a student, he was receptive to guidance. He was motivated by the skepticism he encountered rather than being paralyzed by it. Nakamura and Csikszentmihalyi go on to describe the social conditions and the state of the field that helped the ideas of Linus Pauling be accepted into the "culture" of chemistry.

Research has examined how laypeople and experts view the creative person. For example, Sternberg (1985) found that people's implicit theories of creativity contained such elements as "connects ideas," "sees similarities and differences," "has flexibility," "has aesthetic taste," "is unorthodox," "is motivated," "is inquisitive," and "questions societal norms." He found differences in such characterizations across disciplines. Professors of art placed heavy emphasis on imagination, originality, and a willingness to try out new ideas. Philosophy professors emphasized playing imaginatively with combinations of ideas and creating classifications and systematizations of knowledge different from the conventional. Physics professors were focused on inventiveness, the ability to find order in chaos, and the ability to question basic principles. Business professors emphasized the ability to create and explore new ideas, especially as they related to novel business products or services.

Creative Ideas

Creative ideas provide novel perspectives on phenomena in ways that provide insights not previously recognized. Ideas differ in their degree of creativity, with some ideas being extremely creative while other ideas are only marginally so. Markedly creative ideas have been characterized by Sternberg as “crowd defying.” Sternberg comments on the use of one’s intellect to lead as opposed to “defy the crowd”:

> Some people use their intelligence to please the crowd, others to defy it. The most traditionally intelligent ones hope to lead the crowd not only by accepting the presuppositions of the crowd but also by analyzing next steps in thinking and by reaching those next steps before others do. (2002, p. 376)

By contrast, crowd-defying ideas eschew the presuppositions on which a body of knowledge is based.

Just because an idea is “crowd defying” does not make it useful. Many of Linus Pauling’s ideas that were unrelated to his primary contribution were groundbreaking but proved to be wrong, such as the triple helix structure of DNA and the value of vitamin C for fighting colds. Similarly, “crowd-defying” ideas can meet unexpected resistance, as was the case with Charles Darwin’s theory of evolution and the resistance it engendered not from scientists but from religious groups. Thomas Young’s theory of light was so controversial in 1910 that it was viewed as a “negative contribution,” only later to be recognized as years ahead of its time. The politics of a creative theorizing are complex, and we delve into this subject in more detail in Chapter 13.

The Creative Process

Research on creativity has focused on the creative process itself. For example, Amabile (1983) characterized creativity as involving three core facets: (1) having high motivation to work on the task at hand, (2) having domain-relevant knowledge and abilities to address the task, and (3) having creativity-relevant skills. Creativity-relevant skills include cognitive styles that allow one to cope with complexities and to break one’s “mental set” during problem solving (i.e., to make shifts in one’s chain of thought); the use of heuristics for generating novel ideas; and a work style typified by concentrated effort, an ability to set aside problems, and high energy.

Sternberg and his colleagues (Sternberg, Grigorenko, & Singer, 2004) have developed a theory that conceptualizes creativity as being a function of six resources: (1) intellectual abilities, (2) knowledge, (3) styles of thinking, (4) personality, (5) motivation, and (6) environment. With respect to intellectual abilities, Sternberg stresses the importance of the ability to see problems in new ways and to escape conventional thinking, the ability to discern which ideas are worth pursuing, and the ability to persuade others about one's ideas. In terms of knowledge, Sternberg emphasizes the need to know enough about a field to move it forward, but warns that such knowledge also can result in closed and entrenched perspectives. This is one of the major challenges for creative thought—being knowledgeable about a field, but not letting that knowledge channel thinking too much.

With respect to thinking styles, Sternberg argues that creativity requires a preference for thinking in novel ways as well as the ability to think both globally and locally. To invoke an old cliché, one must not only see the trees, but one must also be able to distinguish the forest from the trees. The personality traits that Sternberg emphasizes for creativity include a willingness to overcome obstacles, a willingness to take sensible risks, a willingness to tolerate ambiguity, and self-confidence and assurance. In terms of motivation, Sternberg emphasizes the importance of task-focused motivation. Numerous studies suggest that people rarely are creative unless they really love what they are doing and focus on the work rather than the rewards that potentially derive from that work. Finally, Sternberg stresses the importance of a supportive environment that rewards creative ideas. Without some environmental support, creativity will be suppressed rather than manifested.

Simonton (1988, 2004) has reviewed the literature on scientific genius and studied the lives of great scientists and concludes that creative ideas often happen randomly, spontaneously, and fortuitously. To be sure, for these chance events to have an impact, the individual also must have exceptional logic and intellect to recognize the underlying connections and take advantage of them. But Simonton's thesis is that chance plays a far larger role in scientific creativity than is often recognized. Simonton also notes that creative events are more likely to occur for scientists who have a strong interest in scientific disciplines outside their chosen specialty and for those who tend to be "mavericks" and to think and do the unexpected.

Yet another perspective on creativity comes from the applied world of advertising. Full-service advertising agencies tend to have four principal departments: an account management department that interfaces with the client/advertiser, a creative department responsible for developing the words and images used to attract our attention and motivate us to behave, a media department that determines where the finished advertising is placed, and a research department that provides various inputs along the way. The backbone of any advertising agency is its creative department.

As is the case in many applied arenas, when they have been successful, senior people in advertising tend to write books describing the approach they used to become successful. One such influential book, *Applied Imagination*, was written by Alex Osborn (1963), the "O" in BBD&O, one of the foremost advertising agencies in the United States. In addition to coining the term *brainstorming* to describe a creative process used by small groups, Osborn identified seven stages that he claims characterize the creative process of both groups and individuals: orientation, preparation, analysis, ideation, incubation, synthesis, and evaluation. At the first two stages, Osborn suggested that one should become familiar with the problem or phenomenon of interest, reading, researching, and learning about its essentials and complexity. In the third and fourth stages, Osborn advises that one should consider the problem or phenomenon from as many different perspectives as possible. In the process of doing so, one develops as many different potential solutions or ideas as possible.

Regardless of whether the idea-generation process takes hours, days, or weeks, at some point, the creative well will seem to run dry. At that point, Osborn encourages one to "sleep on it," that is, to let the unconscious mind take over. Get away from the problem or phenomenon for a few hours, days, or (if you can) weeks. Let things incubate and percolate. It is during this period of incubation, often when one least expects it, that the ideas one has generated synthesize into one's most creative insights. For scientists, Osborn's last stage, evaluation, essentially translates into empirically testing one's creative output.

Deciding to Be Creative

Sternberg (2002) emphasizes the importance of "making the decision to be creative." For creativity to occur in science, it typically is preceded by a personal decision to try to think creatively. Sternberg suggests that social scientists should encourage their students to "decide for creativity" and to inoculate them from the challenges and obstacles that come from making this decision. "Deciding for creativity" does not guarantee creativity, but without such a decision, there is lessened hope for creativity.

Practical Implications for Theory Construction

While there is a sizeable literature on creativity that we cannot comprehensively review here, our brief consideration of this literature drives home several points you should consider. Creative scientific thinking does not require that you revolutionize a field with every idea you generate. Creative ideas cover the gamut from small and incremental to large and revolutionary, with the former being much more typical. Creativity, no matter how big or small, means adding something new, and it typically involves "thinking outside the box." Creative people redefine problems, analyze their ideas, and then try to persuade others of the value of their ideas. They take sensible risks and seek connections between ideas that others do not seek; at least at some level, they realize that existing knowledge can be as much a hindrance as it is a help in generating creative ideas. Creative contributions are the result of the interaction of many factors, only one of which is the act of generating the creative idea itself. If you can't communicate those ideas and get people excited about them, the ideas probably will fall flat. Thus, you must consider communication strategies as well as idea-generation strategies.

In our experience, one of the first steps in generating creative ideas is to start with the proper mindset. Decide to be creative and to think outside the box. Declare to yourself that you are open to mixing the uncommon with the common. Commit to generating ideas without analyzing too much their merits and demerits. You can screen out bad ideas later. Your focus should be one of getting a range of ideas, no matter how odd they initially might seem. Adopt a mindset that you are willing to overcome obstacles, willing to take sensible risks, and willing to tolerate ambiguity. Most of all, be self-confident and have faith in your intellect. There will be features of your environment (including other people) that discourage creative thinking, and it may be difficult to convince others of the merit of your ideas. But we agree with Sternberg that a crucial first step is making the decision to be creative. Do not expect creativity to be a substitute for hard work. Creativity builds upon hard work.

CHOOSING WHAT TO THEORIZE ABOUT

The first step in building a theory is choosing a phenomenon to explain or a question/problem to address. The reasons scientists choose one particular phenomenon or problem to study rather than another are diverse. Some scientists study a phenomenon because they are genuinely interested in it. For example, a scientist might study the mental processes involved in playing chess because he or she finds such phenomena to be intrinsically interesting. Other scientists study a given phenomenon because of its practical or social significance (e.g., reduction of headaches, poverty, globalization). Graduate students frequently study a phenomenon because their advisors study that phenomenon. Some scientists choose to work in areas for which grant funds are available. Other scientists study a phenomenon because people whom they respect also are studying that phenomenon. In many ways, the selection of a phenomenon to study is a personal matter involving the value system of the theorist. In this sense, science is not "value free." Our values and social milieu dictate which phenomena we seek to understand (see Brinberg & McGrath, 1985, for a more extended discussion of the role of values in science).

When thinking about a topic or problem to study or a question to answer, it is helpful to ask "What is interesting about that problem/question?" and "Why is that an important topic?" (Alford, 1998). Be careful about selecting areas that are too broad and abstract. For example, choosing to build a theory about "adolescent risk behavior" involves a construct that is diverse and includes such topics as adolescent drug use, tobacco use, sexual risk taking, delinquency, and alcohol use, to name a few. The explanatory mechanisms might be quite different for these various instantiations of "adolescent risk behavior," and it may be too big a task to tackle theorizing about adolescent risk behavior in general. Literally thousands of studies have been conducted to understand each of the separate risk behaviors mentioned above. We do not want to discourage abstract thinking across instantiations of a construct, and indeed, for the present example several interesting "grand" theories of adolescent risk behavior in general can be found in the literature (e.g., Jessor & Jessor, 1977; Jessor, 1994). However, when choosing phenomena to study, it is advisable to exercise caution in delimiting the scope of a theory. We offer this advice with reservations, because some of the most powerful theories in the social sciences are ones that operate at higher levels of abstraction and thus find applicability in multiple content domains. However, when working at the abstract level, the possibility of obscuring important details and lapsing into vagueness are challenges to confront. We return to this point later in this chapter and in Chapters 11 and 12.

As you think about phenomena to study, you invariably also must think about the population of individuals about whom you will theorize. People differ on many facets, and there are an infinite number of ways in which you can delimit a population. Does the theory you will build apply to infants, toddlers, elementary-age schoolchildren, preadolescents, adolescents, young adults, adults, and/or older adults? Across what dimensions of the larger population do you want your theory to be applicable? At this point, you may not want to delve into the issue of generalizability (the ability to extend your findings to different populations) too much, but you do want a reasonable sense of who you are going to be theorizing about. The initial decision about who to focus on is not "set in stone" as the theory construction process unfolds. Sometimes in thinking more about your concepts and constructs, you will realize that they apply to a more restricted population of individuals than you initially thought. Or, you may realize that they apply to a wider population of individuals than you thought. But at the outset, it usually is best to have a specific population "in mind" as you begin your thoughtful endeavors.

Another strategy for identifying problem areas and questions to focus on is to use the framework of participatory action research (McIntyre, 2007; Reason & Bradbury, 2001, 2007). Participatory action research takes many forms, but it generally involves working directly with entities (e.g., social groups, organizations, communities, towns, cities) to identify the problems with which they are faced, the research that must be conducted to address those problems, and then implementing that research and the solutions suggested by the research in conjunction with members of the entity in question. Participatory action research is not an exotic variant of consultation; rather, it is designed as active coresearch by and for those who are to be helped (Wadsworth, 1998). Although one may decide not to pursue participatory action research in its entirety, certainly the process of interviewing and working with members of the target social entities to identify worthwhile problems on which to work is potentially useful.

It is one thing to try to solve an existing problem or answer an existing question, but creative scientists also identify new problems to solve or new questions to answer. Some of the most influential theorists in the social sciences are individuals who have identified new problem areas to study or who have reframed problems. As examples, in anthropology, John Cole and Eric Wolf (1974) framed issues and questions to define the new field of *political ecology*, which is the study of how political, economic, and social factors affect environmental issues (see also Enzensberger, 1974). Amos Tversky and the Nobel laureate Daniel Kahneman changed the analysis of decision making by bringing to bear the concept of heuristics (i.e., simplified cognitive "rules of thumb") and asking questions about the kinds of heuristics that people use when making decisions (Kahneman & Tversky, 1973; Tversky & Kahneman, 1973, 1974). This framework challenged traditional models of decision making that emphasized rational choice and subjective-expected utility frameworks.

LITERATURE REVIEWS

Perhaps the most often recommended strategy for gaining perspectives on a phenomenon or question/problem is to consult the scientific literature already published on the topic. A comprehensive literature review may serve not only as a useful source of ideas, but is an essential prerequisite for scientific research. Having said this, it may surprise you to learn that some scientists do not always seek out the scientific literature when first theorizing about a topic (Glaser, 1978). The idea is that reading the literature may prematurely narrow your thinking and make it difficult to think "outside the box" in ways that are new and creative. To be sure, if the literature is not consulted initially, it *always* is consulted after the scientist has generated his or her initial ideas. But some scientists prefer to avoid becoming too immersed in a literature when initially thinking about a topic. This is not to say that the topic is approached with a blank slate relative to the existing literature. Usually, the academic training, past readings, and research of the scientist over the years provide him or her with perspectives and knowledge that can't help but be brought to bear. But the intent is to rely more on heuristics and analytic strategies (described below) than on a formal literature review during the initial idea-generation phase.

Having said that, the majority of scientists reject the "hold off reviewing the literature" strategy. Most scientists maintain that reading the literature stimulates ideas that might not otherwise come about. It also provides much-needed focus and clarity before embarking on the theory construction process, and it avoids the possibility of spending time inventing the wheel—only to later find out that someone else already invented that very same wheel. These scientists stress the importance of reading the literature thoroughly, critically, and creatively. In the final analysis, there is no one correct way to go about theory generation, and either approach to the extant literature might work well for you. However, the norm is to delve into the extant literature in depth before building your own theory.

HEURISTICS FOR GENERATING IDEAS

We now turn to specific strategies you can use to think about issues in creative and novel ways. These heuristics are ones that we or other scientists have found useful, but they may or may not resonate with you for the particular task on which you are working. We personally do not use every heuristic, and we find that some work better for some problems than for others. The heuristics are only a start to building the tools in your theory construction toolbox. In later chapters we develop additional ways of thinking that will augment or complement the strategies described here. If you use these heuristics, they may help you think about a phenomenon in ways that are different from how you might otherwise proceed. But, if a simple list of heuristics was all it took to generate creative and innovative ideas, then creative ideas would abound in science. It is not simple!

Idea Generation and Grounded/Emergent Theorizing

As discussed in Chapter 10, some social scientists argue for the use of emergent methodologies when constructing theories. The essence of these approaches is that theorists should set aside, as much as possible, their preconceptions when studying a phenomenon and let ideas about the concepts and the relationships between concepts emerge as they embed themselves in the surroundings and contexts of the groups under study. Theory is thereby "grounded in" and emerges from such data, with the data typically being qualitative in nature.

Some grounded theorists reject the idea of doing anything more in theory construction than letting concepts and relationships emerge in the context of careful observation in natural settings. To be sure, creativity and insight are involved in the framing of issues to be investigated and in recognizing subtleties in the world that might not be obvious to the casual observer. But the emphasis is on setting aside preconceptions, staying close to the data, and letting theory emerge "from the ground up."

Other grounded theorists are more open to imposing novel ways of construing and thinking about observed events in their natural settings. These theorists try to think about their observations and their field notes in creative ways, such as by invoking metaphors, drawing analogies, relating observations to other constructs that form the bases of their own interpretive systems, and so on. For these grounded theorists, one or more of the heuristics discussed in this section may resonate as useful.

Twenty-Six Heuristics

This section presents 26 heuristics for thinking about phenomena or questions in a way that might give you new insights and ideas. It assumes you have already identified a phenomenon or problem area to study.

1. *Analyze your own experiences.* This strategy involves thinking about your own experiences and reflecting on what factors have influenced the outcome variable or the phenomenon you are trying to explain based on them. For example, if the outcome variable is social anxiety, you might think about situations in which you have been nervous or anxious and identify the features of those situations relative to other situations that caused you to be nervous. You might think about situations where you overcame your social anxiety and reflect on how you did so. You might think about periods in your life when you tended to be more socially anxious and periods when you tended to be less socially anxious and reflect on what aspects of your life differentiate those periods. The idea is to carefully analyze your own experiences and see if this helps you think about

what you are trying to explain in new and creative ways. For grounded theorists, thinking about your field notes and observations relative to your own personal experiences might give you perspectives on how to frame arguments and propositions about the observations you have made.

2. *Use case studies*. This strategy for idea generation involves analyzing a single case (individual, family, group, or organization) in detail and generating ideas based on that case study. It might take the form of formally interviewing a person, family, group, or members of an organization in depth, and/or researching archival data and other existing data sources about the person, family, group, or organization. Case studies range from labor-intensive enterprises that take years to complete, to simpler gatherings of information about a single person or event. There is a substantial literature on methods for conducting case studies (Eisenhardt, 1989; Cosentino, 2007; Ellet, 2007), and readers are encouraged to consult this literature to gain a better sense of the many forms of case studies. Sigmund Freud used this approach to develop facets of his highly influential psychoanalytic theory. Examples of case studies leading to important theoretical insights in the social sciences abound (see, e.g., Spatig, Parrott, Kusimo, Carter, & Keyes, 2001; Turner, 1970). You should consider the possibility of pursuing case studies as a method for idea generation.

3. *Collect practitioners' rules of thumb*. This strategy involves interviewing or researching experts in an area who are actively dealing with the phenomenon you want to study (Mayo & LaFrance, 1980). For example, to formulate a theory of the best way to counsel children about grief-related experiences, you might interview professionals who do such counseling to obtain their perspectives on how best to do this. Or you might interview scientists who study such professionals and/or who conduct research on the matter. Personal interviews (in person, over the phone, through e-mails) with such experts often can yield a richer account of a phenomenon than what you would garner from reviewing the formal scientific literature on the topic. The collected rules of thumb might then serve as the basis for a more formal theory (see Hill, 2006; Kaplan, 1964). In his work on social influence, Cialdini (2003) collected practitioners' rules of thumb for exerting social influence by working in the professions of, and interviewing, insurance and car salespeople, charitable fundraisers, and a host of other professionals who worked in occupations that rely on social influence. Based, in part, on these collected rules of thumb of influence, Cialdini formulated novel theoretical perspectives on persuasion and social influence.

4. *Use role playing*. This heuristic involves putting yourself in the place of another and anticipating how this person might think or behave with respect to the outcome variable. For example, put yourself in the place of another family member, a friend, or someone you know whose experiences might be relevant to the phenomenon you are trying to understand. Think about how they would approach matters. Put yourself in the place of another scientist who has different training from your own and imagine how he or she would conceptualize or think about the problem or explain the phenomenon. Role playing is widely used in organizational training and other applied contexts to foster new levels of understanding and insight (El-Shamy, 2005). Here we are suggesting a

technique of "mental role playing," as you intellectually take the role of others and think about their reactions and perspectives. For grounded theorists, thinking about your field notes relative to the perspective you invoke by role playing relevant others might give you perspectives on how to conceptualize and frame arguments about the observations you have made.

5. *Conduct a thought experiment*. Another tool for generating ideas—one often used by Albert Einstein—is to conduct a thought experiment (Ackoff, 1991; Folger & Turillo, 1999; Lave & March, 1975; Watzlawick, 1976). These are hypothetical experiments or studies that are conducted "in the mind," as if you have collected data and then imagine the results. In this strategy, you think about relevant variables or scenarios and then think about the effects of different variations or manipulations of them, as if you were running an experiment—but instead doing all this in your head. You imagine the results and think about what implications they have for explaining your phenomenon. In other words, to understand fact, think fiction.

A substantial literature exists on thought experiments, primarily in philosophy, political science, economics, and physics. Sorensen (1998) examined thought experiments as used by theorists in a variety of disciplines and concluded that they were central in the theory construction process (see also Nersessian, 2002). Tetlock and Belkin (1996) explored the use of a particular type of thought experiment, namely the use of *counterfactuals*, to focus on "what-might-have-been" questions. More technically, counterfactuals are subjunctive conditionals in which the antecedents of an event are assumed to be known but for purposes of argument supposed to be false. For example, you might ponder the counterfactual, "If the United States had not dropped atomic bombs on Japan, then the Japanese would have surrendered at about roughly the same time that they did." Counterfactual thought experiments can sensitize us to theoretical possibilities we might otherwise have ignored, had we not pursued counterfactual thinking.

As an example in the social sciences, an investigator might be interested in the consequences of having an abortion for adolescents in the United States. A counterfactual thought experiment would be to think about the implications of abortion for adolescents if abortions were illegal. Thinking through this counterfactual circumstance might suggest consequences of having an abortion that the theorist may not have thought of initially. The economist Richard Fogel used a counterfactual to explore the economic impact of the emerging railroads on American economic growth (Fogel, 1964). He explored the counterfactual of what the American economy would be like had there been no railroads. Interestingly, his detailed exploration of the counterfactual suggested that the railroads did not have a significant impact on the American economy and, more generally, that no single innovation was vital for economic growth in the 19th century.

Try to frame the phenomena with which you are working in terms of different thought experiments and counterfactuals. Chapter 6, on clarifying relationships, provides detailed applications of thought experiments for social science phenomena, so we defer further discussion of this technique to that chapter.

6. *Engage in participant observation*. Participant observation involves observing oth-

ers in the situations you want to study while you actively participate in those situations. For example, you might live in the community and work with a group of factory workers to get a better idea of factors that influence worker attitudes toward labor unions. Like case studies, participant observation can range from labor-intensive efforts lasting years to more short-lived efforts (Hume & Mulcock, 2004; Reason & Bradbury, 2007). Participant observation is a component of many ethnographic analyses, a dominant approach for advancing theory in anthropology. It has been used to generate rich theoretical analyses of a wide range of important phenomena. For example, the anthropologist Margaret Mead (2001) spent months in Samoa using participant observation to study the lives of teenagers as she explored cultural influences on adolescence. She dispelled the then-common view that adolescence is a time of "stress and turmoil" as adolescents search for an adult identity. Participant observation is discussed in more detail in Chapter 10. There we see that some social scientists believe that pursuing theory construction without such qualitative data is folly.

7. *Analyze paradoxical incidents.* Sometimes it is useful to isolate and analyze a paradoxical situation. For example, suppose you are interested in the occurrence of unintended pregnancies. As you reflect on this phenomenon, you recognize that some unmarried individuals who are sexually active, do not want to become pregnant, and have positive attitudes toward birth control nonetheless fail to use any form of contraception. This appears to be paradoxical behavior. What could be operating to explain this paradox?

This particular paradox was explored in a qualitative study by Edin and Kefalas (2005), who studied low-income, inner-city young women, mostly Latinas and African Americans, living in the United States. Edin and Kefalas found that about two-thirds of the pregnancies in the women they studied were not planned. Rather, pregnancy occurred by chance, whenever a woman happened to stop using birth control. Contraception often was used at the beginning of a relationship but was discontinued when a relationship "reached another level." Edin and Kefalas report that both men and women in their study interpreted a woman wanting to have a baby by a man as a high form of social praise for that man. Thus, nonuse of birth control was linked to an expression of respect and social praise.

The women in the study tended to see children as giving meaning to a woman's life. Women who placed careers over motherhood were seen as selfish. Childbearing and marriage were not seen as decisions that "go together." This disconnection, however, did not reflect a disinterest in marriage. To the contrary, women held high standards for the men they would be willing to marry. They hoped that the fathers of their children would rise to the occasion and be worthy of being a life-long partner, but they did not count on it. Also, marriage, the women said, was reserved for those men and women who had "made it" economically and were worthy of it. Women thought that both partners should be economically "set" prior to marriage. Indeed, women expressed an aversion to economic dependence on a male. To be worthy of marriage, the women said, couples must have demonstrated relational maturity by withstanding hard times together. This can take years to attain. As a result of such dynamics, women were less likely to marry.

The result of the work by Edin and Kefalas was to provide the beginnings of a theory of why many sexually active, unmarried, inner-city women paradoxically do not seek to become pregnant but fail to use birth control. The women are willing to leave having a baby to chance, and see the nonuse of birth control as the sign of a good relationship that can yield the positives associated with childbearing.

As with the other heuristics, analyzing paradoxical incidents can be pursued mentally using the strategies of imagined role playing, by analyzing your own experiences, by using thought experiments, or it can take the form of a qualitative study.

8. *Engage in imaging.* Most of the time, we think about problems in verbal terms. It is as if we have an internal conversation with ourselves. We subvocally think through our thoughts and ideas. The current heuristic involves setting aside linguistic-based thoughts in favor of visualizing relevant situations and behaviors. For example, in thinking about causes of social anxiety, try to imagine yourself at a party with few people you know. Try to visualize this situation as graphically as possible and in as much detail as possible. Visualize the setting and the people who are in that setting. Now start playing out the interactions you have with other people. But do not just verbally note these interactions to yourself. Try to imagine them happening, as if you were watching a movie. The general idea is to draw upon imaging and the cues that imaging stimulates as you think about a phenomenon. This strategy tends to engage the right hemisphere of the brain, as opposed to linguistic thoughts, which engage the left hemisphere (Gregory, 1997). New insights may result, accordingly.

Imagistic thinking and visualization have been found to be common in thought experiments conducted by creative scientists, as they explore in their minds the use of old schema in new situations outside their normal area of application, especially when spatial reasoning is required (Gooding, 1992; Nersessian, 2008). Visual intelligence has been the subject of considerable empirical research (Arnheim, 2004; Barry, 1997; Hoffman, 2000), and its role in the generation of scientific ideas is also being explored (e.g., Clement, 2006, 2008; Nersessian, 2008).

9. *Use analogies and metaphors.* This heuristic involves applying the logic of another problem area to the logic of the area of interest or drawing upon a metaphor. For example, several theories of memory use the metaphor of a “storage bin,” where information is placed in long-term memory for later access (Bodenhausen & Lambert, 2003). Long-term memory is seen as consisting of thousands of such bins, which are organized in a complex fashion. Pieces of information relevant to a bin are stacked on top of each other within a bin, with the most recently processed information at the top. Thus, the most recently placed information in a bin is more easily accessed when a person retrieves information from that bin. Of course, the physiology of the brain does not contain physical storage bins, but the metaphor promotes a way of thinking about memory and making predictions about information retrieval processes. Metaphorical thinking thus provides a mechanism for building a theory of memory.

As another example, Randall (2007) used a compost heap as a metaphor for autobiographical memory, arguing that it comes closer than more commonly used computer analogies to capturing the dynamics of memory across the lifespan and how memory changes with aging. Randall describes the parallels between composts and such memory-related concepts as encoding, storage, and retrieval and derives an organic model of memory to better understand the psychology of aging.

The process of using analogies can be complex. For example, Clement (2008) gave scientists challenging problems in their areas of study and collected think-aloud protocols as they attempted to solve them. Clemens found that scientists tended to struggle with the analogies they generated, often resorting to “bridge analogies” to link the problem with the original analogy. Imagistic simulations and visualization also were often used to think through and explore the application of analogies. The scientists Clement studied engaged in cycles of analogical thought in which they generated, critiqued, and modified series of analogies. They showed great persistence in pursuing analogies. Charles Darwin is known to have worked through a single analogy over a period of years in developing his theory of evolution (Millman & Smith, 1997).

Clement (1988) described four processes scientists use to apply analogies to solve a problem. First, the scientist generates the analogy. Second, the scientist establishes the validity of the analogy in relation to the original problem. Third, the scientist seeks to understand the analogous case. Finally, the scientist applies the findings to the original problem. Clement (1988) identified three strategies that scientists use to generate analogies at the first step. First, they might generate an analogy from a well-known principle. This step involves recognizing that the original problem situation is an example of a well-established principle. Second, the scientist might create an analogy by modifying the original problem situation, thereby changing features of it that were assumed to be fixed. Finally, the scientist generates an analogy through association in memory, whereby the scientist is “reminded” of, or recalls, an analogous case in memory.

The use of metaphors in social science theories is common. For example, as noted above, metaphors and analogies are widely used in the analysis of memory and memory processes. In sociology, the classic work of Erving Goffman (1959, 1967) viewed everyday interactions between people using an analogy of actors following scripts, imposing their own interpretations onto the scripts, and employing occasional ad libs. In management, Cornelissen (2004) uses the metaphor of “organizations as theatre” to develop basic tenets of an organizational theory.

10. *Reframe the problem in terms of the opposite.* This heuristic involves reversing the focus of your thought to a focal opposite. If you are trying to understand the reasons why some people are highly loyal to a brand when purchasing a particular product, it might help instead to think about why people are not brand loyal. The work of William McGuire is a good example that combines the use of thinking about opposites in conjunction with the metaphorical thinking discussed in the previous section (McGuire, 1968). McGuire was a noted theorist in the field of attitude change who approached persuasion-related phenomena thinking not about how to influence people but instead about how to make people resistant to attitude change. McGuire used biological immunization principles as a metaphor for inducing resistance to persuasive communications. He noted that people are inoculated against many diseases by introducing small amounts of a contaminating virus into the body. As a result, the body builds up antibod-

ies to resist the virus, thus preventing the full-blown occurrence of the disease should higher levels of viral exposure be experienced later. To make people more resistant to attitude change, McGuire provided them with short, challenging, counterattitudinal messages that would make them think through their attitude in more detail and organize their defenses to counterarguments. These "immunization" challenges, McGuire reasoned, should not be compelling enough to change the attitude, but rather should be just strong enough to mobilize counterarguments and defense mechanisms to combat future persuasive attempts. McGuire used the immunization analogy to effectively develop a complex theoretical framework for the analysis of persuasion.

11. *Apply deviant case analysis.* Sometimes a certain individual or group of individuals will stand out, with respect to a phenomenon, as being different from the rest of the crowd. This heuristic focuses on "deviant" cases in the attempt to explain why they are deviant. For example, although most adolescents who have good relationships with their parents do not use drugs, others who have good relationships with their parents do use drugs. Why? What can explain the behavior of these "deviant" cases?

An example of deviant case analysis is a research program that one of the authors developed (Jaccard) examining the relationship between perceived knowledge and actual knowledge in different knowledge domains (Jaccard, Dodge, & Guilamo-Ramos, 2005; Radecki & Jaccard, 1995). For instance, Jaccard and colleagues studied people's perceptions of how knowledgeable they thought they were about nutrition and compared this with how knowledgeable they actually were about nutrition as assessed on knowledge tests. Jaccard observed a nontrivial number of people who significantly overestimated their knowledge levels; that is, they perceived themselves as being quite knowledgeable when the knowledge tests indicated they had very little knowledge about the topic at hand. Rather than ignore these "deviant cases" relative to the more typical case of those showing correspondence between perceived and actual knowledge, Jaccard explored why these individuals so overestimated their knowledge levels. In doing so, he was able to articulate and build on broader theories of the bases of perceived knowledge.

12. *Change the scale.* Mills (1959) suggests imagining extreme changes as a method of stimulating thinking: If something is small, imagine it to be enormous and ask "What difference might that make?" If something is pervasive, think about what things would be like if it were not. An example of this heuristic is the research on globalization in the fields of geography, anthropology, political science, and economics. Theorists in these disciplines think about how processes operating at the global level might be mirrored at the local level, as well as how processes operating at the local level might be generalized or altered if they are moved to the global level (Goldin, 2009; Robins, Pattison, & Woodstock, 2005). For example, if we think about the effects and challenges of unionizing workers at the local level, what would happen if we tried to unionize workers at the global level? How might thinking in such terms alter our theories and thinking about unionization at the local level?

13. *Focus on processes or focus on variables.* The dominant tradition in the social sciences is to think in terms of variables. An alternative approach is to think in terms of processes—that is, sets of activities that unfold over time to produce change, to main-

tain equilibrium in a system, or to get from event *A* to event *B*. Abbott and Alexander (2004) describe the implications of thinking about crime and criminals in process terms. Instead of thinking about a criminal as someone who commits a crime, a process analysis views the act of becoming a criminal as consisting of the sequenced actions of getting caught, detained, held, charged, convicted, and sentenced. The often observed inverse relationship between social class and becoming a "criminal" might be due, in part, to the fact that lower-class individuals are more likely to make it through this process-based sequence than middle- or upper-class individuals. Process-oriented models have existed in the social sciences for decades, but they are not as popular as variable-centered approaches (Gilbert & Abbott, 2005; Cederman, 2005). If you tend to think primarily in terms of variables, try thinking in terms of dynamic processes. Conversely, if you tend to think primarily in terms of processes, try thinking in terms of variables.

One strategy for invoking process perspectives is to change nouns into verbs (Weick, 1979). The use of this heuristic is most evident in the field of organizational studies, where process theories change such vocabulary as *order* to *ordering* (Cooper & Law, 1995), *being* to *becoming* (Chia & Tsoukas, 2002), and *knowledge* to *knowing* (Cook & Brown, 1999) so as to invoke process-oriented explanations. The premise of process-based analyses is that organizations are in continual flux and that variable-centered theories are limited because they ignore this flux, capturing only momentary "snapshots" of organizations at a single point in time (De Cock & Sharp, 2007; Sturdy & Grey, 2003).

Process-oriented perspectives are central to the field of processual anthropology, which was articulated by Turner (1967, 1970) in his temporal analyses of rituals. Turner argued that rites of passage are marked by a progression through three stages: (a) separation, when a person becomes detached from a fixed point in the social structure; (b) marginality, when the person is in an ambiguous state, no longer in the old state but not yet having reached the new one; and (c) aggregation, when the person enters a new stable state with its own rights and obligations. Turner characterizes each stage as well as factors that influence and condition movement from one stage to another. This process-oriented view of phenomena contrasts with more structuralist interpretations of rituals.

For those who already think in terms of processes, a way of generating new ideas might be to think instead in terms of variables. Abbott (2004) recommends a heuristic of "stopping the clock." The idea is to "freeze" the process at a given point in time and then describe the system in detail at the frozen moment. By "stopping the clock," you broaden the context and apply other heuristics we have discussed to that particular point in the process.

Try analyzing your phenomenon as a dynamic process that fluctuates over time and that entails moving from event *A* to event *B*. How do entities get from one point to another? We describe process-oriented frameworks in more detail in Chapter 9 on simulations, Chapter 10 on grounded theory, and Chapter 11 in the context of broader systems of thought.

14. *Consider abstractions or specific instances.* Using instantiation principles discussed in Chapter 5, or the opposite of instantiation, abstraction, think about the

phenomena at different levels of abstraction. For example, when thinking about what influences people's attitudes toward a political candidate, think about what influences attitudes in general and then apply this to your analysis of attitudes toward political candidates. If you are thinking about what influences people's attitudes in the abstract, then think about what influences a person's attitude toward a political candidate (as well as other specific attitudes) and try to generalize from this position to attitudes in general.

Arie Kruglanski (2004), a noted social psychologist, emphasizes the importance of abstraction in theory construction. Kruglanski states that a key strategy for formulating abstract principles or constructs is to seek commonalities among phenomena and to be skeptical about distinctions. He argues that the surface manifestations of different instances of the same phenomena can be misleading and that different concrete phenomena often are driven by the same underlying principle. The focus should be kept on isolating the gist of the phenomena and identifying what absolutely must be known about it. For example, attitude change, conformity, majority–minority influence, and social power all have a social influence component that may allow you to build a general theory of social influence by looking for communalities in each of these phenomena. Rather than focusing on domain-specific constructs or processes, strive to isolate underlying principles that generalize across domains. Having said that, Kruglanski also urges caution in the use of abstraction. When abstracting to more general levels, it is possible to obscure important distinctions that should be made. The key is to always ask yourself what is gained and what is lost by moving across different levels of abstraction.

The sociologist Robert Alford (1998), like Kruglanski, views the theory construction process as one of constantly moving back and forth between reflective musings about the implications of abstract concepts to concrete analyses of specific observations. Theory is developed by thinking about concrete instantiations of concepts and then abstracting upward to more general constructs that allow us to make theoretical propositions that generalize across many content domains.

15. *Make the opposite assumption*. Take an explicit assumption and recast it to its opposite. If a phenomenon is assumed to be stable, think of it as unstable. If two variables are assumed to be related, what would happen if they were unrelated? If two phenomena coexist, what would happen if they could not coexist? If *X* is assumed to cause *Y*, consider the possibility that *Y* causes *X*. This heuristic is similar to the ones discussed earlier, on framing a problem in terms of its opposite and "changing the scale." However, it is distinct in subtle ways.

As an example, self-esteem is often thought of as being a relatively stable construct in which some people have chronically low self-esteem and others have chronically high self-esteem. Suppose we think instead about self-esteem as being transitory rather than stable—that is, self-esteem fluctuates from one day to the next. This line of thinking might lead us to suggest a two-component theory of self-esteem. The first component is relatively stable across time and reflects the overall level of self-esteem that an individual experiences across time. The second component is transitory in nature and reflects the daily variation in self-esteem in relation to the stable component. On some days, the individual's self-esteem might be lower than what is typical for that person, and on other days, the individual's self-esteem might be higher than what is typical for that person. What factors impact the stable component of self-esteem, and what factors impact the transitory component of self-esteem? How might the two components differentially affect important outcome variables, such as school performance or engaging in risk behaviors? How might one further develop this two-component conceptualization of self-esteem? By thinking about this stable construct in unstable terms, new insights result. Could the same type of stable–transitory analysis be invoked for other constructs that are assumed to be stable over time? For example, could ethnic identity be viewed as having the two components described above?

As another example, it is typically thought that the quality of parent–adolescent relationships influences drug use by adolescents, with better relationships being associated with lower probabilities of drug use. But what if we reverse the causal direction? Is it possible instead that an adolescent's use of drugs negatively impacts the relationship between the adolescent and his or her parents? As another example, it is commonly accepted that adolescents who are religious are less likely to engage in problem behaviors—that is, the religiosity has a protective function. Could it be instead that adolescents who start to engage in risk behaviors (e.g., alcohol or drug use) become less religious? Perhaps lowered religiosity is not a risk factor but instead is just a by-product of the adolescent's engaging in problem behaviors and then rejecting his or her (religious) upbringing in the process. As a final example, it is often assumed that parental childrearing strategies impact child behavior, but it is just as plausible to assert that child behaviors influence parental childrearing strategies.

Abbott and Alexander (2004) provide several additional examples of this heuristic in their book on theory construction. For example, it is commonly assumed that college educates students. But suppose instead we think about all the ways in which college prevents education (e.g., boring classes, emphasizing rote memorization). As we make opposite assumptions and try to marshal support for them, we can gain new insights into the phenomena we study.

16. *Apply the continual why and what.* Given an outcome or dependent variable, ask yourself, "Why do some people do this but other people do not?" or "Why do some people have more of this but other people do not?" For each answer you give, repeat this question again. For example, if the outcome variable is school performance, you might ask, "Why do some people do well in school, whereas others do not?" The answer might be, "because some people are more intelligent and more motivated." Then ask "Why are some people more intelligent?" and "Why are some people more motivated?" The answer to the former might be "because of the way they were raised." This last answer is vague, so you might ask "What do you mean by that?" After clarifying this statement, you might ask yourself, "Why were they raised like that?" As you continually probe successive "why" questions for their answers and ask, "What do you mean by that?," you might gain new insights into the original question you posed.

This strategy was used by Darryl Bem (1970) in his analysis of the psychological bases of beliefs. Bem focused his analysis on a belief that people might hold, such as "Smoking cigarettes causes cancer," and asked people why they believed this to be the case. When given an answer, he would ask them why they believed the answer was true. He would then pose the "Why do you think that is true?" question to the new answer that was given. As he continually pushed beliefs back to their origins, he found that all beliefs rested on one or both of two foundations, namely, (a) the belief that something is true because an authority figure or expert says it is true, and/or (b) the belief that something is true because it was directly experienced (or someone else directly experienced it) and "one's senses do not lie." How do you know, for example, that the sun is the center of the solar system? The answer probably is "because I was taught this in school by my science teacher"—which is invoking an authority figure. The continual "why" heuristic led Bem to derive an interesting theory of the bases of cognitions (see Bem, 1970).

17. *Consult your grandmother—and prove her wrong.* McGuire (1997) coined the term "bubba psychology" to refer to the idea that much of what social scientists "discover" in their research is so banal or obvious that it could have been told to them by their grandmother ("bubba"). Take the obvious and think about how it could be wrong. Try to turn bubba social science on its head. Or, as a variant, extend an obvious "bubba fact" to situations where you might be surprised by their implications. For example, the bubba fact that "being in a good mood increases your life satisfaction" might be extended to a subtler idea that "finding a dollar on the street increases life satisfaction." If you find a dollar, doing so may put you in a good mood. If being in a good mood positively influences your current feelings of life satisfaction, then it follows that finding a dollar on the street should raise your life satisfaction (at least temporarily). People might find the idea that "coming upon a dollar on the street improves life satisfaction" to be more intriguing than the idea that "being in a good mood increases life satisfaction." By extending the obvious to the nonobvious, an interesting theoretical point might be made.

18. *Push an established finding to the extremes.* Take a well-established relationship or finding and consider the extremes of it. For example, it is often held that more eye contact with a person will increase his or her liking of you. But what happens if you make constant eye contact? Too little or too much eye contact may be bad, but an intermediate level of eye contact may be just right. Parents who are affectionate with their children tend to have happier children. But what if a parent is constantly affectionate? Some scientists have suggested that too much of anything eventually backfires or starts to produce opposite effects. What happens at the extremes of your phenomenon?

19. *Read biographies and literature, and be a well-rounded media consumer.* Much of social science focuses on people, and an excellent resource for the lives of people is published biographies and autobiographies. These can be a rich source of ideas about many facets of human behavior. Fiction and nonfiction works are filled with insightful analyses, as are movies and other forms of media. The humanities and arts, even though they do not to rely on formal scientific methods, are an invaluable source of ideas and insights. Take advantage of the wealth of knowledge in these different areas.

One of the authors (Jaccard) attended the University of California at Berkeley as an undergraduate and recalls his first meeting in the office of the noted sociologist Erving Goffman. While waiting for Professor Goffman to finish a phone call, Jaccard casually noted the books on his shelves. They included the typical textbooks of a sociological academic, but they also included popular magazines typically seen at grocery store checkouts, cookbooks, books on fashion, popular self-help books, and other books representing a wide swath of American culture. It was only after taking several courses from Professor Goffman that Jaccard came to appreciate that these resources were a core source of Goffman's ideas about human interaction and American society.

20. *Identify remote and shared/differentiating associates.* Think about the phenomenon of interest to you and try to identify as many causes and consequences of it as you can. Be expansive, listing as many plausible ones as possible. The idea is to create a "free association" scenario, in which the list you generate includes "remote associates," that is, things that are unlikely for people to think of and that might spark a creative insight. Relatedly, think of people you know or have heard of who perform the behavior or phenomenon you are trying to explain and people you know who do not perform the behavior or the phenomenon. Then make a list of the qualities, attributes, and characteristics of each type of person. On what qualities and attributes do they differ and on which ones are they the same? What do the similarities and differences tell you about the phenomenon?

21. *Shift the unit of analysis.* When theorizing, we often do so by describing and explaining behavior at the level of individuals. For example, we may want to know why some people are wealthy but other people are poor. Or, we may want to know why some people can cope well with a debilitating disease but other people cannot. However, there also are theories that focus on "units of analysis" other than individuals, such as a couple, a family/household, a small group, an organization, or even countries within a world system. In such cases, the outcome variable is not the behavior of the individual but instead it is the behavior of the "unit." For example, in the context of the HIV epidemic, we might focus on understanding why *couples* engage in unprotected intercourse rather than trying to explain why *individuals* engage in unprotected intercourse. Or, in the field of organizational studies, we might focus on understanding factors that influence the productivity of *organizations* rather than the productivity of *individuals* within organizations.

The noted sociologist James Coleman (1994) emphasized the importance of analyzing the functioning of social systems rather than individuals. Note that this is not the same as studying the impact of social systems on individuals, as would be the case, for example, when exploring the impact of work environments on life satisfaction. Instead, the focus is on the behavior of the unit per se—that is, the particular social system in question. According to Coleman (1994), theories of the functioning of social systems can focus either on forces or explanatory constructs outside the system that impact and shape how the system of interest operates or, alternatively, on forces and processes within the system, that is, the component parts of the system, that impact or shape how that system operates. For example, the productivity of an organization may be impacted by the broader economy in which it functions as well as the internalization of an organizational "culture of productivity" on the part of the individuals who work within the organization.

If you tend to work with phenomena at the individual level, try shifting the "unit of analysis," if plausible, and consider the phenomena from this new "unit" level. As an example, research has explored the effects of alcohol on adolescent tendencies to engage in unprotected sex, and almost all of this research uses individuals as the unit of analysis. However, unprotected sex is a couple behavior, so you could shift the analysis from the individual level to the couple level by focusing on factors that influence whether a given *couple* engages in unprotected sex. One way of analyzing couple behavior is to do so in terms of the different attitudes and characteristics that each couple member brings to the relationship. For example, for condom use, the male member has a behavioral orientation toward using condoms, a set of beliefs about using condoms, certain risk-taking orientations, certain alcohol consumption patterns, and so on. This is also true of the female member of the couple. That is, she also brings to bear a behavioral orientation toward using condoms, a set of beliefs about condoms, certain risk orientations, and certain alcohol consumption patterns. The couple's behavior is some function of the meshing of these two sets of variables, one set from the male and one set from the female.

Four relational models describe possible ways in which the attributes of each couple member combine to impact couple behavior. Consider the case of the effects of alcohol use on couple sexual behavior. The first model, termed the *female influence model*, states that alcohol use impacts the sexual behavior of a couple, but that the primary influence on such behavior is the level of alcohol consumption by the female member of the dyad as opposed to the male member of the dyad. That is, alcohol consumption by the male does not impact the couple behavior but female alcohol consumption does. The second model, termed the *male influence model*, occurs when male drinking behaviors, rather than female drinking behaviors, influence the couple's sexual activity. The third model, termed the *shared influence model*, conceptualizes the drinking behaviors of both the male and female partners as being independent determinants of couple-based sexual risk behavior. The final model involves more complex couple dynamics and is termed the *interactional influence model*. This model is a variant of the shared influence model in that it accounts for risky sexual couple behavior using the drinking behaviors of both partners, but it allows for configural influence. In addition to the independent effects of the drinking behaviors of both couple members, the interactional influence model posits an interaction effect between couple member drinking behavior, such that as female drinking increases, male drinking behavior becomes more strongly related to risky sexual behavior.

This example frames the analysis around the "components" of the system. However, analyses also could be pursued using unit-level variables as well, such as couple intimacy, relationship length, couple communication, and couple bargaining/negotiation strategies. For interesting perspectives on couple-level analyses, see Kenny, Kashy, and Cook (2006) and the references cited therein. Note that by shifting the unit of analysis from individuals to couples, very different perspectives on the effects of alcohol on unprotected sex have been gained.

Zaheer, Albert, and Zaheer (1999) extend the concept of units of analysis to temporal dimensions. They suggest that thinking about process and change phenomena from the perspective of different units of time can yield new theoretical insights and ideas. For example, if you are building a theory of the dynamics of life satisfaction, what concepts and variables might you focus on if you seek to understand changes in life satisfaction on a daily basis, on a monthly basis, on a yearly basis, or on the basis of decades? By thinking in these different units of time, new ideas and perspectives on life satisfaction may result.

22. *Shift the level of analysis*. Social scientists focus on explanations at different levels of analysis. One way of characterizing levels of analysis is in terms of proximal versus distal determinants. Proximal determinants are the more immediate determinants of behavior, and distal determinants are variables that influence behavior but do so through the more immediate determinants. Some theorists explain phenomena using more proximal determinants, whereas others explain phenomena using more distal determinants. We often gain new insights into a phenomenon by shifting the level of analysis we are pursuing, either by moving from proximal to distal analysis or from distal to proximal analysis.

As an example, Jaccard (2009) has presented a framework for thinking about behavior at four different levels. At the first (proximal) level, the explanatory framework asserts a simple proposition: A person's behavior is influenced by his or her intention to perform the behavior. If people intend to do something, they usually will do it, and if they do not intend to do something, they usually will not do it. For example, if a person intends to get tested for HIV, he or she probably will do so. If the person does not intend to get tested for HIV, he or she probably will not do so. In actuality, the relationship between behavioral intentions and behavior is complex—that is, people do not always do what they intend to do. The theory elaborates why this is the case. Factors that affect the intention–behavior relationship include environmental constraints that impede intended behavioral performance, lack of relevant knowledge and skills to perform the intended behavior, forgetting to perform the intended behavior, and the operation of habit and automatic processes, among others.

At the second level of analysis, the near-proximal level, the theory addresses why some people intend to perform a behavior and other people do not. Five classes of variables that impact a person's intention or decision to perform a behavior are the focus of the theory: (a) what the person sees as the advantages and disadvantages of performing the behavior, (b) the normative pressures that the person experiences to perform the behavior, (c) the perceived social image implications of performing the behavior, (d) the person's emotional and affective reactions to performing the behavior, and (e) the person's self-perceived ability to successfully perform the behavior (i.e., perceptions of self-efficacy). For example, what does a person see as the advantages and disadvantages of being tested for HIV; what social pressures are operating for the person to be tested for HIV; how will being tested for HIV affect the person's image he or she conveys to others; what emotional reactions does the person have to being tested for HIV; and, what obstacles does the person see to being tested for HIV?

At the next level of explanation are near-distal determinants. These are more general variables that do not reference the target behavior, per se, but can shape the proximal and near-proximal determinants of behavior. They include such constructs as (a) personality variables; (b) general values, goals, aspirations, and attitudes; (c) mental-health-related variables (e.g., depression, anxiety, stress); and (d) variables related to alcohol and drug use. For example, a person may miss a scheduled HIV test because he or she is depressed or hung over from drinking too much.

The most distal level of analysis focuses on the broader contexts in which behavior occurs: the family context, the peer context, the school context, the work context, the provider context, the religious context, the neighborhood context, the media context, the government/policy context, and the cultural context (for various ethnic groups).

Jaccard (2009) encourages theorists who seek to explain behavior to think about that behavior at the different levels of analysis. Do people intend to perform the behavior in question, and if so, what factors get in the way of their carrying out their intentions? Why do some people intend to perform the behavior whereas others do not? How do these people differ in their perceived advantages and disadvantages of performing the behavior, the normative pressures that are operating, the perceived image implications of performing the behavior, their emotional reactions to performing the behavior, and their perceived ability to perform the behavior? How are all these variables shaped by their personalities, their broader goals and aspirations, their general attitudes, and other such lifestyle variables? Finally, how is all this influenced by the broader contexts in which they live, including the family context, the peer context, the school context, the work context, the religious context, the neighborhood context, the media context, the government/policy context, and the cultural context?

Most social scientists tend to theorize at only one level of analysis, either at the level of the more proximal determinants of behavior or at the level of more distal determinants of behavior. This focus is perfectly reasonable. However, it might help to gain insights into the phenomenon you are studying by occasionally shifting your thought process to think about explanatory constructs either at more proximal levels or more distal levels. By thinking about phenomena at these different levels, insights might be gained into the kinds of variables you should focus on at the original level of analysis. For example, even though you might be interested in understanding the political ideologies of voters and how these influence voting choices, it might be helpful to shift to another level of analysis and ask how different contexts (e.g., the media, work, neighborhood) shape the political ideologies of individuals. Engaging in such an activity might suggest new dimensions of ideology you had not considered or new ways of thinking about ideology that might help you better explain the relationship between ideology and voting behavior.

23. *Use both explanations rather than one or the other.* In the area of impression formation, a robust finding is the occurrence of primacy effects. Information about a person that is presented first tends to have a larger impact on impressions than information that is presented later (Anderson, 1965, 1991). One explanation for this focuses on "change-in-meaning" processes, whereby the initial information is thought to color or change the meaning of the later information so that it is interpreted to be consistent with the initial information. A second explanation is that the initial information is remembered more easily than later information. More easily recalled information is more influential, hence the primacy effect. A third explanation is that people discount later information

that is contrary to the initial information they receive. Research has attempted to choose between these explanations (change in meaning vs. accessibility vs. discounting), with somewhat mixed success (Anderson & Hubert, 1963; Anderson, 1965, 1991). Instead of viewing explanations as mutually exclusive, consider instead the possibility that all of the explanations are operating. Thus, in the impression formation example, perhaps there is some change in meaning operating, perhaps there also is some differential recall operating, and perhaps there is some discounting operating, all converging to produce primacy effects. A theory might be devised whereby each of these processes has an "importance weight" reflecting the influence of the process in the formation of impressions. We might then theorize about how the relative magnitude of these importance weights varies across different individuals and across different situations. This approach yields a different theoretical perspective from that of trying to choose the one "correct" explanation. (For an elaboration of this orientation, see Poole & Van de Ven, 1989.)

A variant of this heuristic is to invoke the operation of multiple processes rather than a single process when thinking about a phenomenon. Rather than viewing a phenomenon as being influenced by one process or the other, allow both processes to operate. This strategy is well illustrated by the many "dual-process" theories that are popular in psychology. Dual-process models take many forms, but the idea is to specify two alternative or complementary processing modes and then build a theory around those processes. For example, Smith, Zarate, and Branscombe (1987) suggest a dual-process model for accessing attitudes: one process that calls a previously formed and previously stored attitude from memory and the other process that accesses a rule from memory to use to form an attitude toward an object to which one is newly exposed. Smith and DeCoster (2000) suggest that people possess two memory systems for storing cognitions: one system that slowly learns general regularities and the other system that forms quick representations of novel events or individual episodes. Petty and Cacioppo (1986) suggest a dual-process model of persuasion based on the systematic processing of a persuasive message (i.e., thinking about the validity of the arguments contained in a message) and the heuristic processing of a persuasive message (i.e., thinking about the source of a message and the characteristics of the source, such as the source's trustworthiness and credibility). Kowalski (2006) offers a dual-process model of thought involving intuitive thinking, which is automatic, effortless, and largely subconscious, and deliberative thinking, which is controlled, effortful, and mostly conscious. Sierra and Hyman (2006) offer a dual-process model of cheating intentions, one based on cognitive thought and the other based on anticipated emotions. Although dual-process models are most popular in the areas of decision making, memory, and information processing, they also appear in other areas of inquiry. Perhaps the phenomena you are thinking about can be conceptualized in the form of a dual-process framework.

24. *Capitalize on methodological and technological innovations*. Technology is changing rapidly, leading to new methodological tools for social scientists. These advances are opening up new areas for theoretical inquiry and insight. For example, in the past, linking neuroscience findings to human behavior was limited to the exploration of animals removed from their ecological contexts, to observations of patients who suffered

BOX 4.1. The Nacirema

A strategy for generating ideas used by anthropologists is the careful and systematic observation of others in their cultural contexts. By examining cultures in as objective and systematic a way as possible, insights into behavior become apparent that otherwise would remain hidden. This is illustrated nicely in the following edited account of the Nacirema and their body rituals, as described by the noted anthropologist Horace Miner (1956):

> Professor Linton first brought the ritual of the Nacirema to the attention of anthropologists twenty years ago, but the culture of this people is still very poorly understood. They are a North American group living in the territory between the Canadian Cree, the Yaqui and Tarahumare of Mexico, and the Carib and Arawak of the Antilles. Little is known of their origin, although tradition states that they came from the east.
>
> Nacirema culture is characterized by a highly developed market economy that has evolved in a rich natural habitat. While much of the people's time is devoted to economic pursuits, a large part of the fruits of these labors and a considerable portion of the day are spent in ritual activity. The focus of this activity is the human body, the appearance and health of which loom as a dominant concern in the ethos of the people. While such a concern is certainly not unusual, its ceremonial aspects and associated philosophy are unique.
>
> The fundamental belief underlying the whole system appears to be that the human body is ugly and that its natural tendency is to debility and disease. Incarcerated in such a body, a person's only hope is to avert these characteristics through the use of ritual and ceremony. Every household has one or more shrines devoted to this purpose. The more powerful individuals in the society have several shrines in their houses and, in fact, the opulence of a house is often referred to in terms of the number of such ritual centers it possesses. Most houses are of wattle and daub construction, but the shrine rooms of the wealthier people are walled with stone. Poorer families imitate the rich by applying pottery plaques to their shrine walls.
>
> While each family has at least one such shrine, the rituals associated with it are not family ceremonies but are private and secret. The rites are normally only discussed with children, and then only during the period when they are being initiated into these mysteries. I was able, however, to establish sufficient rapport with the natives to examine these shrines and to have the rituals described to me. The focal point of the shrine is a box or chest that is built into the wall. In this chest are kept the many charms and magical potions without which no native believes he or she could live. These preparations are secured from a variety of specialized practitioners. The most powerful of these are the medicine men, whose assistance must be rewarded with substantial gifts. However, the medicine men do not provide the curative potions for their clients, only decide what the ingredients should be and then write them down in an ancient and secret language. This writing is understood only by the medicine men and by the herbalists who, for another gift, provide the required charm.
>
> The charm is not disposed of after it has served its purpose, but is placed in the charm box of the household shrine. As these magical materials are specific for certain ills and the real or imagined maladies of the people are many, the charm box is usually full to overflowing. The magical packets are so numerous that people forget

cont.

what their purposes were and fear to use them again. While the natives are very vague on this point, we can only assume that the idea in retaining all the old magical materials is that their presence in the charm box, before which the body rituals are conducted, will in some way protect the worshipper.

Beneath the charm box is a small font. Each day every member of the family, in succession, enters the shrine room, bows his or her head before the charm box, mingles different sorts of holy water in the font, and proceeds with a brief rite of ablution. The holy waters are secured from the water temple of the community, where the priests conduct elaborate ceremonies to make the liquid ritually pure.

In the hierarchy of magical practitioners, and below the medicine men in prestige, are specialists whose designation is best translated as "holy-mouth-men." The Nacirema have an almost pathological horror of, and fascination with, the mouth, the condition of which is believed to have a supernatural influence on all social relationships. Were it not for the rituals of the mouth, they believe that their teeth would fall out, their gums bleed, their jaws shrink, their friends desert them, and their lovers reject them. They also believe that a strong relationship exists between oral and moral characteristics. For example, there is a ritual ablution of the mouth for children that is supposed to improve their moral fiber. The daily body ritual performed by everyone includes a mouth rite. Despite the fact that these people are so punctilious about care of the mouth, this rite involves a practice that strikes the uninitiated stranger as revolting. It was reported to me that the ritual consists of inserting a small bundle of hog hairs into the mouth, along with certain magical powders, and then moving the bundle in a highly formalized series of gestures.

In addition to the private mouth rite, the people seek out a holy-mouth-man once or twice a year. These practitioners have an impressive set of paraphernalia, consisting of a variety of augers, awls, probes, and prods. The use of these items in the exorcism of the evils of the mouth involves almost unbelievable ritual torture of the client. The holy-mouth-man opens the client's mouth and, using the above mentioned tools, enlarges any holes that decay may have created in the teeth. Magical materials are put into these holes. If there are no naturally occurring holes in the teeth, large sections of one or more teeth are gouged out so that the supernatural substance can be applied. In the client's view, the purpose of these ministrations is to arrest decay and to draw friends. The extremely sacred and traditional character of the rite is evident in the fact that the natives return to the holy-mouth-men year after year, despite the fact that their teeth continue to decay.

Miner goes on to describe numerous other features of the Nacirema culture. Now, spell the name "Nacirema" backwards and reread Miner's account of the Nacirema. Does it mean something different to you now? This exercise drives home the point that what can seem normal and everyday from one perspective can be seen as unusual, even bizarre, from another perspective. The tendency to see things from our own cultural viewpoint is called an *ethnocentric* perspective. It shapes our thinking in predetermined ways. Creative individuals can often break out of their ethnocentric constraints.

trauma, to disorders of localized areas of the brain, and to postmortem examinations. Recent technological advances now permit measures of electrophysiological recording, functional brain imaging, and neurochemical assessments during ongoing human behavior, permitting intriguing interfaces between social science and neuroscience. As an example, Steinberg (2008) suggests that the brain's socioemotional system develops at a much faster rate than the cognitive control system, leading adolescents toward increased reward seeking at a time when control mechanisms are underdeveloped. He hypothesizes that this pattern of brain development accounts for risk-taking increases during adolescence followed by risk-taking declines during young adulthood, as brain development surrounding control processes "catches up" with those surrounding emotional responses. Steinberg is breaking new theoretical ground in developmental science by taking advantage of technology-based methodological advances in neuroscience.

As another example, Freud popularized the notion of the unconscious in psychology, but it quickly fell into scientific disrepute when satisfactory measures of unconscious phenomena failed to materialize. Without such measures, unconscious motives could be introduced post hoc to explain anything because there was no way of subjecting unconscious explanations to empirical tests. Recently, psychologists and sociologists have developed methods that purportedly allow for the measurement of unconscious attitudes (Bassili & Brown, 2005; Blanton & Jaccard, 2008). These methods ask individuals to view stimuli on computer screens and to classify them as fast as they can into different categories. Based on how long it takes them to classify stimuli (measured in milliseconds), inferences are made about the attitudes they have toward the stimuli they are classifying. These methods have opened up new theoretical accounts of human behavior, as both conscious and unconscious phenomena are incorporated into conceptual frameworks.

Keeping abreast of new technologies and methodological innovations is a way of possibly defining new questions and bringing to bear new constructs to understand the phenomena you study.

25. *Focus on your emotions*. Emotions are seen by many as having no place in theory development, but when studying human behavior, emotions might be tapped for purposes of generating ideas. It is a common practice in anthropology, for example, to record field notes either as an outside observer or in the context of participant observation. Field notes not only involve recording one's observations, but they also involve recording notes about the emotions one is experiencing while making the observations. In traditional anthropology, such emotion notes were considered to be "warnings" about potential bias in one's observations, such as when one's perceptions of others might be colored by the relationships that had been formed with those others in the context of participant observation. However, more recently, some groups of anthropologists have viewed emotion notes as a source of meaning and interpretation in their own right that may help them formulate theory and gain perspectives on the phenomenon being studied (e.g., Kelinman & Copp, 1993; Cylwik, 2001). The anthropologist Rosaldo (1993) used his grief over the death of his wife as a way to understand the intensity of emotions experienced by a group of headhunters, the Ilongot of the Philippines, about headhunt-

ing. Rosaldo argues that by attempting to eliminate personal emotions from observations, traditional ethnographies create distortions and misinterpretations of descriptions, thereby undermining explanation.

26. *What pushes your intellectual hot button?* Items in the media or discussions with others may evoke in you a sense of disbelief or at least disagreement. Any time you find yourself saying "That can't be right," ask yourself if it is something worth pursuing. As a junior Assistant Professor discussing the value of advertising with a consumer advocate, one of the authors (Jacoby) found himself disagreeing with the advocate's proposition that "if there is going to be advertising, it should be limited to providing as much concrete information to the prospective consumer as possible." Incredulous, Jacoby contended that, if it were all attended to, too much information likely would discombobulate the consumer, making it more difficult to separate the wheat from the chaff, thereby leading to poorer decision making. The advocate scoffed at the idea. This led Jacoby to conduct research to test his hypothesis. After Jacoby published a series of studies that confirmed the hypothesis, advertisers and public policymakers began thinking differently about how information could best be communicated to consumers. As an example, instead of requiring cigarette manufacturers to list the complete set of 18 health consequences it wanted on cigarette packages, the Federal Trade Commission, citing this stream of "information overload" research, called for cigarette manufacturers to list only three of these consequences at a time, periodically rotating the consequences to be listed. What have you heard or seen that elicits your disagreement and makes you say "That can't be right"? Is it something worth studying?

Other heuristics could be mentioned, and we will develop more of them in subsequent chapters. The present list is simply a start. A compendium of other approaches to stimulating creativity can be found in the volumes by Stein (1974, 1975) and McGuire (e.g., 1997). Nersessian (2008) emphasizes the fact that scientists rarely rely on one heuristic or one cognitive process for generating ideas, but rather notes that an idea might result from the complex interplay of multiple heuristics, such as mixing of the use of analogies, thought experiments, and imaging. Try thinking up some heuristics of your own. How can you start to "think differently?" How can you view things from perspectives you are not accustomed to? How can you combine heuristics? Try creating your own list of ways for stimulating creativity.

When the Focus Is on Basic Mental or Biological Processes

Some social scientists focus their theorizing on core processes within the human mind, such as attention, perception, or comprehension, whereas others focus on basic biological processes, such as neural correlates of thought or electrical activity in the brain that is associated with different emotions. Can the heuristics and categories of variables described above be used by such theorists? Yes. This is not to say that every heuristic or every class of variable will be of use, but if you make a sincere effort at thinking about phenomena from the perspective of one or more of the heuristics, there is a good chance that new insights will emerge. The use of heuristics—such as interviewing experts, conducting thought experiments, analyzing paradoxical incidents, imaging, using analogies and metaphors, reframing problems in terms of their opposite, using deviant case analysis, focusing on processes, considering abstractions or specific instances, making the opposite assumption, applying the continual why and why not questions, pushing an established finding to the extremes, using remote associates, shifting the level of analysis, and using multiple rather than either–or explanations—all can be applied to the analysis of basic mental, biological, organizational, or sociological processes.

SCIENTISTS ON SCIENTIFIC THEORIZING

Articles have been written by leading social scientists on the strategies they use for constructing scientific theories. In this section we review some of these strategies, focusing on those that complement or augment the heuristics already discussed. One or more of these approaches may resonate with you as you think about your phenomena.

Robert Wyer is an influential social psychologist who studies the phenomenon of information processing. His theory of information processing emphasizes seven processing stages or activities in which individuals engage when processing information: (1) attending to information when it is encountered, (2) interpreting and organizing the information relative to preexisting concepts that are stored in memory, (3) construing the implications of the information for already acquired knowledge about relevant people and events, (4) storing the information in memory, (5) retrieving the information from memory at later times, (6) integrating the retrieved information to make a subjective judgment, and (7) translating this judgment into an overt response. In building a scientific theory about information processing, Wyer (2004) notes that he often compartmentalizes these processes, building a mini-theory for each one separately. These mini-theories are then aggregated to construct his broader theory of information processing. Compartmentalization thus can assist in building theories of complex phenomena. In other words, take it one step at a time.

Wyer (2004) states that he always searches for alternative explanations of a phenomenon. What different assumptions, suppositions, or perspectives might lead to the same phenomenon occurring? When presented with an explanation for a phenomenon, Wyer's immediate reaction is to generate an alternative explanation for it. Then he carefully analyzes the different explanations, deciding either to integrate them or to competitively test them against one another in an experiment.

Wyer emphasizes the importance of bringing fresh perspectives to theory construction. He expresses reservations about reading existing literature at the idea-generation phase. The writers of research reports, he notes, can be very effective at conveying their logic and thereby channel the reader's thinking to that of the writers' thinking. He states that he often reads the first few paragraphs of an article and then skips to the Method and Results section so that he can generate his own explanations of the findings. Then he goes back and reads what the authors have to say.

Wyer notes that theories often are metaphorical in character, and he encourages the use of metaphors when theorizing. Wyer was one of the developers of the "bin" theory of memory described above, and his work is typified by the use of rich and creative metaphors.

John Cacioppo (2004) discussed the use of reductionism in his theorizing. *Reductionism* is the attempt by scientists to identify, break apart, or reduce nature into its natural constituents. Identifying the core ingredients of a phenomenon reveals insights into that phenomenon. Although some have criticized reductionism as making the simplistic assumption that "the whole is equal to the sum of its parts," Cacioppo argues that it provides points of entry into complex systems. The idea is not just to describe the smaller parts but to develop a better understanding of the complex system. Reductionism is a strategy for doing so.

William McGuire (2004) argues for the importance of expressing theories using six different modalities: (1) verbal, (2) abstract symbolic, (3) pictorial (or graphic), (4) tabular, (5) descriptive statistical, and (6) inferential statistical. McGuire argues that expressing a theory via multiple modalities can help the theorist grasp the theory better and can increase the likelihood of noticing the implications of a theory and its similarities to and differences from other formulations. As examples, Jacoby (2002) worked with the traditional stimulus–organism–response ($S \rightarrow O \rightarrow R$) model in psychology, which conceptualizes stimuli in the environment as impinging on a person (or, more generally, an organism), who then formulates a response to that stimulus. He asked, "What if the traditional $S \rightarrow O \rightarrow R$ model is not depicted linearly but instead is conceptualized as an overlapping Venn diagram?" Representing the framework as a Venn diagram led Jacoby to propose new ways of representing how stimuli, responses, and organismic factors interact, with the result being a richer, seven-sector conceptualization of the traditional $S \rightarrow O \rightarrow R$ model. In another realm, Langley (1999) describes several creative visual mapping strategies that can be used for characterizing process-oriented theories, and McGuire (2004) provides several additional examples of expressing theoretical propositions in each modality.

Howard Becker, an internationally recognized sociologist, wrote a book called *Tricks of the Trade* (2003) that is filled with heuristics for generating ideas. We describe some of them here, focusing mainly on those that augment the material already presented. One heuristic Becker describes is a *next-step heuristic*. Suppose a person performs a behavior. What are the next steps that are available to him or her? If you graduate from high school, what are the next steps you can take in your life? How does this action (graduating from high school) enhance or constrain the next steps that you can take? If you are late to work, what are your next steps? If you purchase a product (e.g., a car), what are your next steps? Becker argues that theory construction is analogous to a form of storytelling that is constrained by the demands of being logical and consistent with known facts. Stories lead the reader from one step to the next. What events lead up to your "primary event" (i.e., your outcome) and what events follow it? As you think about matters, keep applying a "next step" line of thinking: Every time you specify the next step, ask yourself what is the next step that can or should be taken after that.

Another strategy Becker discusses is one of "building a machine to maintain the status quo." Suppose you are interested in academic achievement. What would you need to do to keep the current levels of learning in schools exactly where they are at? What steps would you need to take to ensure that neither any improvements nor any decrements were made in student performance? How could you prevent teachers from doing a better job without also making them do a worse job? How could you keep parents from doing a better job without also making them do a worse job with respect to their children and school? If you had to "build a machine" or "design a system" to maintain the status quo exactly as it is, what parts would the machine have, how would they function, and how would they interconnect? As you answer these questions, you may gain insights into how to improve academic achievement.

Social scientists generally try to specify *why* a person performs a particular behavior. As another heuristic, Becker suggests shifting away from this traditional *why* question to focus instead on *where* someone performs the behavior. Think through this strategy in depth. If you interview someone for a case study analysis, probe in detail the question of where the behavior occurs. Such a shift will help you focus on settings and the importance of settings in influencing behavior. For example, someone may buy one brand of beer for personal consumption at home, but often select another brand of beer when drinking with his buddies. Another question shift is to ask *how* instead of *why*. Instead of asking why people smoke marijuana, ask how someone came to start smoking marijuana. What were the steps leading up to it? How did the smoker get to the circumstances he or she is in today? A final question shift can be to ask *when* does someone perform the behavior rather than why. By carefully thinking about "where," "how," and "when," in addition to "why" people do what they do, more insights into the phenomena in which you are interested may result.

Yet another strategy that Becker suggests is to doubt everything that anyone in a position of authority or who supposedly is an expert tells you. By accepting nothing as a given, by being a doubter about everything, by looking for alternative interpretations, new ideas can suggest themselves.

Abbott and Alexander (2004) have written a useful book describing heuristics to stimulate thinking about social science phenomena, many of which we have already discussed. Additional heuristics they note include (1) questioning scientific propositions that others take as givens (similar to the suggestion of Becker, noted above); (2) reconceptualizing an action or outcome as not being due to an actor but instead being due to a device or a circumstance, such as when Ralph Nader reconceptualized injuries from car accidents as the result of poorly designed cars that were inherently unsafe rather than speeding drivers; (3) setting conditions, whereby one specifies those conditions under which a relationship holds and those conditions under which the relationship does not hold (a topic we take up in detail in Chapter 6), such as stress having a larger negative impact on well-being for people with poor coping skills as opposed to people with good coping skills; (4) adding a dimension, which involves identifying conceptual confounds and controlling for them to see if a relationship still remains (e.g., for the proposition "Women are less likely to pursue mathematics than men": Would this be true if one controlled for or held constant parental encouragement? Would it be true if one held constant mathematical abilities?); (5) splitting, which involves making distinctions that one does not typically make (e.g., as in Epstein's [1983] book *Women in Law*, in which the author argues that traditional studies and characterizations of lawyers apply to male but not female lawyers, or in Joseph Carlin's (1962) book *Lawyers on Their Own*, in which he makes the same argument but for lawyers in small, solo practices); and (6) redefining constructs in novel ways, such as when West and Zimmerman (1987) defined gender not as a variable or as a role but instead as a set of social actions that included making certain gestures and invoking certain symbols in certain contexts, all with the intent of identifying oneself as being gendered.

Finally, Root-Bernstein and Root-Bernstein (1999) identified 13 core thinking tools that creative geniuses use as they approach the process of idea generation. They identified these tools based on a careful analysis of the writings and lives of highly creative individuals in a wide range of disciplines encompassing the arts, the sciences, and the humanities. The tools include (1) observing, (2) imaging, (3) recognizing patterns, (4) pattern forming, (5) analogizing, (6) abstracting, (7) body thinking, (8) empathizing, (9) dimensional thinking, (10) thought playing, (11) transforming, (12) synthesizing, and (13) modeling. Many of these strategies map onto the heuristics we have already discussed. Details of each are summarized in Root-Bernstein and Root-Bernstein (1999).

SUMMARY AND CONCLUDING COMMENTS

In this chapter we have provided ways of thinking about phenomena that may help you think of new concepts or constructs or consider relationships between concepts and constructs in novel ways. This chapter is just a beginning; the chapters that follow will augment your theoretical toolbox in additional ways. It is not enough for you to passively read about the heuristics we have described in this chapter and then expect a major, creative idea to hit you over the head (unless you are an incarnation of Richard Feynman). Nor should you give a heuristic a superficial try and then reject it as being simplistic or irrelevant. Force yourself to explore each heuristic in depth—at different times and for different phenomena. Maybe you will ultimately decide it is not useful to pursue, but the idea is to stretch your thinking and to try new avenues of thought.

Theoretical advancements in the social sciences can range from the small and incremental to the large and revolutionary, as we've noted. A central process in theory construction is idea generation. When trying to explain something, we want to create new and novel ideas and perspectives that will build on existing knowledge. This is just as true for "small" ideas as it is for large, "revolutionary" ideas. The key to theory construction is to generate numerous ideas from different vantage points and then to screen those ideas in terms of which ones are worthwhile and which ones are not. Once an initial "cut" has been made, the ideas must be subjected to more rigorous explications and analysis, as described in Chapters 5 and 6.

Creative people tend to be independent nonconformists who have drive and resiliency and are willing to take risks. They typically have a burning curiosity and tend to trust their intuition. The prototype of a creative person differs by profession. Creative ideas provide insights not previously recognized. Ideas differ in their degrees of creativity, with some ideas being extremely creative whereas others are only marginally so. As an emerging scientist, you need to "decide to be creative."

As you approach theory construction, you will first identify and frame a problem or question to study and then use heuristics and ways of thinking that help you "think outside the box." We presented numerous heuristics and ways of thinking, some of which may resonate better with you than others. These included:

- Analyzing your own experiences
- Conducting case studies
- Collecting practitioners' rules of thumb
- Role playing
- Conducting thought experiments
- Using participant observation
- Analyzing paradoxical incidents
- Imaging
- Using analogies and metaphors
- Reframing problems in terms of their opposite
- Using deviant case analysis
- Changing the scale
- Focusing on processes
- Considering abstractions or specific instances
- Making the opposite assumption
- Applying the continual why and why not questions
- Proving your grandmother wrong
- Pushing an established finding to the extremes
- Reading biographies and literature
- Using remote associates
- Shifting the unit of analysis
- Shifting the level of analysis
- Using multiple explanations rather than either–or explanations
- Taking advantage of methodological and technological innovations
- Focusing on your emotions
- Relying on your intellectual "hot button"

In addition to these heuristics, you might develop some of your own. Given the decision to be creative, you now have some of the initial tools for theory construction.

The material covered in Chapters 5 and 6 will help you refine your concepts and ideas, and the material in Chapters 7 through 11 will help you organize and use the above in more systematic ways. The creative process does not follow a set progression, so we must jump around a bit until you are exposed to all of the material in the remainder of the book. But the above is a start.

SUGGESTED READINGS

Abbott, A., & Alexander, J. C. (2004). *Methods of discovery.* New York: Norton.—A book describing heuristics for thinking about social science phenomena.

Becker, H. (2003). *Tricks of the trade: How to think about your research while you're doing it.* Chicago: University of Chicago Press.—A book containing some useful tips for generating research ideas.

Jacoby, J. (2002). Stimulus–organism–response reconsidered: An evolutionary step in modeling (consumer) behavior. *Journal of Consumer Psychology, 12*(1), 51–57.—A description of a revised stimulus–organism–response model that relies heavily on imaging, analogies, and metaphors.

McGuire, W. J. (1997). Creative hypothesis generating in psychology: Some useful heuristics. *Annual Review of Psychology, 48*, 1–30.—An insightful discussion of 49 strategies for generating research ideas.

Osborn, A. F. (1963). *Applied imagination* (3rd ed.). New York: Scribner's.—A book describing how to encourage the creative process.

Root-Bernstein, R. S., & Root-Bernstein, M. (1999). *Sparks of genius: The thirteen thinking tools of the world's most creative people.* New York: Mariner.—A discussion of key thought processes used by leading thinkers throughout time.

Simonton, D. K. (1988). *Scientific genius.* New York: Cambridge University Press.—A study of creativity in science.

Simonton, D. K. (2004). *Creativity in science: Chance, logic, genius, and zeitgeist.* New York: Cambridge University Press.—A study of creativity in science.

Sternberg, R., Grigorenko, E., & Singer, J. (2004). *Creativity: From potential to realization.* Washington, DC: American Psychological Association.—A book describing features of the creative process.

Wicker, A. (1985). Getting out of our conceptual ruts: Strategies for expanding conceptual frameworks. *American Psychologist, 40*, 1094–1103.—A discussion of heuristics for generating research ideas.

KEY TERMS

creativity (p. 40)

deciding to be creative (p. 44)

analogies/metaphors (p. 52)

abstraction (p. 55)

participatory action research (p. 46)
grounded theory (p. 48)
case studies (p. 49)
thought experiments (p. 50)
counterfactuals (p. 50)
participant observation (p. 50)
unit of analysis (p. 59)
proximal determinants (p. 61)
distal determinants (p. 61)
dual-process models (p. 63)
reductionism (p. 69)

EXERCISES

Exercises to Reinforce Concepts

1. Discuss the role of idea generation and idea screening in theory construction.
2. What are the key characteristics of the creative person?
3. According to Nakamura and Csikszentmihalyi, what are the three system facets that influence the incorporation of novelty into scientific culture?
4. What are the three core factors that Amabile says underlie creativity?
5. What does it mean to “decide to be creative”?
6. What issues and criteria should you consider when choosing a topic area to study? Elaborate on each of them.
7. What are the advantages and disadvantages of doing a literature review before you start thinking about a phenomenon?
8. What are the four processes that Clement describes as characterizing how scientists apply analogies to solve a problem? Describe each.
9. What are the four levels of analysis characterized by Jaccard? Describe each and give examples.
10. Describe the heuristics for idea generation that most resonate with you and discuss why you find them to be the most useful.

Exercises to Apply Concepts

1. Pick a phenomenon of interest and then generate a set of ideas about factors that influence it, how it functions, or how it influences other factors using one or more of the heuristics discussed in this chapter. Identify the heuristics you used.
2. Create a heuristic for generating ideas that was not discussed in this chapter.

5

Focusing Concepts

People see the world not as it is, but as they are.

—ALBERT LEE

When formulating a theory, researchers usually begin with some phenomenon that they want to understand. The phenomena can be diverse, ranging from overt, observable behaviors (e.g., smoking cigarettes, purchasing a product) to hypothetical concepts that are not directly observable (e.g., depression, suicidal ideation). The phenomena may focus on entities (schools), groups of individuals (females, Latinos), processes (automatic processing of stimuli), systems (immune system responses to stress), or time (developmental changes), to name a few. We discussed in Chapter 4 (and discuss in later chapters) issues surrounding the choice of concepts to include in one's theoretical system. Here, we focus on strategies for specifying and refining conceptual definitions for those concepts that one decides to include in the theoretical system.

We begin by describing the process of instantiation—a method for making abstract concepts more concrete—leading us to a formal characterization of the process of specifying conceptual definitions. Next we consider the problem of surplus meaning, followed by a discussion of practical strategies for formulating clear definitions of constructs. We then discuss multidimensional approaches to constructs and, in turn, strategies that social scientists use to create taxonomies and variables for use in their theories. Finally, we conclude with a historical note on the concept of operationism as a means of defining constructs.

THE PROCESS OF INSTANTIATION

When formulating a theory, scientists frequently deal with abstract concepts. In fact, one way in which theories differ is in terms of the abstractness of the concepts used in the framework. For example, one theorist might develop a theory of how a person's attitudes are related to his or her behavior. Such a theory contains two concepts, "attitudes" and "behavior," which are quite abstract. A second theorist might develop a theory about how attitudes toward going to college are related to whether or not a high school student attends college. In this case, the same basic concepts are involved ("attitude" and "behavior"), but they are more focused: The concepts now concern a specific type of attitude (the attitude toward going to college) and a specific type of behavior (going to college). In the process of theorizing, scientists usually try to strike a balance between being too specific (and hence, having a theory that is narrow in scope) and being too abstract (to the point where the concepts become "fuzzy" and unmanageable). The process used to refine fuzzy concepts is that of instantiation. *Instantiation* is a deliberate process that involves specifying concrete instances of abstract concepts in order to help clarify their meaning. It is fundamental to science and a crucial process for refining initial theoretical ideas.

Instantiation plays another important role in science. In addition to clarifying concepts at a theoretical level, instantiation is a bridge between the conceptual and the empirical realms where the validity of a theoretical expression is subjected to empirical testing. Instantiation helps ensure that a theory is testable. Given the theoretical expression "Attitudes toward college influences whether or not a person goes to college," the scientist must devise a strategy for testing whether this statement has any value as a guide to understanding matters. To accomplish this, the scientist might decide to select a group of high school students, measure their attitudes toward attending college, and then later determine whether or not they go to college. Although this strategy may appear straightforward, the scientist pursuing this strategy would quickly be confronted with innumerable questions in the process of implementing it. Should the students be high school seniors only, or should juniors be included? When assessing attitudes, should one make distinctions between attitudes toward community colleges (i.e., 2-year colleges), 4-year colleges, and universities? What about trade schools? How long should the researcher wait to find out if the student has gone to college? One year? Two years? And so on. Each of these questions requires that the scientist focus the concepts, isolating specific instances of them, and then using these instances for testing the validity of the more general theoretical expression. The process of instantiation accomplishes this and is thus an important aspect of designing empirical tests of a theory.

Most scientists use the term *hypothesis* to refer to the specific empirically based instances that are used to test a more general theoretical expression. For example, based on the theory that "attitudes toward college influence college attendance," a scientist might propose a hypothesis for a study that "the high school seniors in this study will be more likely to attend a 4-year university if they have positive attitudes toward attending college." Although this hypothesis is more concrete, it is still somewhat ambiguous and, in practice, the scientist would have to be even more specific in the delineation of his or her "hypothesis." Scientists are in disagreement about the exact meaning of the term *hypothesis*. We refer to the process of specifying a concrete instance of a theoretical expression for purposes of an empirical test as that of stating a hypothesis. Thus, a *hypothesis* is a statement that (1) is derived from a theoretical expression, (2) is more concrete than the originating theoretical expression, and (3) is tied to the empirical realm in such a way as to permit a test of the originating expression. Like theoretical expressions, hypotheses also consist of concepts and relationships. However, the latter are more limited in scope, more specific, and expressed in testable (or near testable) form. A hypothesis is sometimes called a *focal problem statement* and is a direct result of the process of instantiation.

At some point, almost all theories must be subjected to the instantiation process, either to make the definition of "fuzzy" concepts clearer and more communicable or to derive a direct test of the theory. A theoretical expression usually contains a universe of potential meanings. If the universe of meanings is large, the theoretical expression runs the risk of being vague and ambiguous. Consider the following theoretical expression: "Older people are more knowledgeable than younger people." This expression posits a relationship between two concepts: age and knowledgeability. But what do we mean by each of the two concepts? Does "age" refer to chronological age, developmental age (i.e., one's level of maturation), or psychological age ("You're as old as you feel")? To simplify matters, let us assume that we are referring to chronological age. The concept of "knowledgeability" also is ambiguous. A person could be knowledgeable about a myriad of different things, including music, films, television programs, local happenings, world events, geography, aircraft and aviation, his or her family, his or her job, cooking utensils, restaurants, religious events, clothing, consumer products that are generally available, computers, celebrities, photography, and the breeding of tropical fish. It would take an encyclopedia just to list all the possibilities. We need to be more specific about the type of knowledge to which we are referring.

The Nature of Conceptual Definitions

Instantiation involves generating specific conceptual instances so as to make abstract concepts more concrete; it requires posing and answering questions designed to clarify and explicate the meanings of concepts. Nunnally (1967, p. 89) refers to this process as "outlining constructs" or "stating what one means by the use of particular words." The outcomes of this process are termed *conceptual definitions*, which represent clear and concise definitions of one's concepts. Not only do abstract constructs have to be defined as precisely as possible, but even seemingly obvious concepts frequently require explication. Consider a researcher specializing in the design of retail outlets who is responsible for conducting a study to determine where to erect a new shopping center. The researcher may begin by asking people in the target area what they want most in a shopping mall and learn that "convenience" is highly regarded. The researcher is not, however, quite sure what this means, since, in response to further probing, it is discovered that the same word means many things to many people—including enough places to park one's car, proximity of the shopping center to other places that the individual frequents, and the distance from the individual's home. The researcher decides to explore the issue of distance further. However, the question "How far away is the *XYZ* shopping center?" produces five different replies from five different people: two subway stops;

three traffic lights; 10 minutes; 3 miles; and quite far. Faced with such responses, a theorist must think carefully about the essence of a concept and then clearly communicate that essence to other scientists.

SHARED MEANING, SURPLUS MEANING, AND NOMOLOGICAL NETWORKS

Researchers sometimes disagree as to what constitutes the essence of the phenomenon under question. Thus, various conceptual definitions may be provided for the same concept. For example, Jacoby and Chestnut (1977) noted there were more than 50 definitions of "brand loyalty" in the field of consumer behavior. Given that each investigator makes his or her conceptual definition explicit (by providing clearly articulated and precisely defined propositions), specific points of agreement and disagreement can be identified. The points of agreement may then be assumed to represent the essential (i.e., agreed-upon) core of the concept. Researchers refer to this as *shared meaning*, which can be contrasted with the remainder, which is termed *surplus meaning*. In science, it is better to have concepts that are dominated by shared meaning and that do not have too much surplus meaning.

Some examples will make explicit the relevant issues. Fishbein and Ajzen (1975) reviewed research on attitudes and found more than 500 definitions in the scientific literature. Even a cursory review of this literature made evident that quite different constructs were being invoked to represent "attitudes." Fishbein and Ajzen argued that three distinct constructs were being called "attitude" and that lumping these constructs together under a common rubric was counterproductive to scientific advancement. Although all definitions of attitude recognized that an attitude is directed toward an "object" (e.g., an attitude toward "majoring in psychology," an attitude toward "going to church," an attitude toward "Republicans"), the definitions varied in other respects. One set of constructs used to represent attitudes were people's perceived associations between the attitude object and some other attribute, characteristic, trait, or concept. For example, a person might believe (rightly or wrongly) that Republicans are conservative, that Republicans oppose abortion, that Republicans are from relatively wealthy families, and that Republicans oppose big government. Fishbein and Ajzen referred to such perceived associations as "beliefs." Another set of constructs used to represent attitudes were general feelings of favorableness–unfavorableness or how positive or negative one felt about the attitude object. Fishbein and Ajzen noted that such affective feelings were an element of most definitions of attitude and, further, that all the standard attitude scaling methods were designed to measure such affective feelings. Because this represented "shared meaning" among the many conceptions of attitude, Fishbein and Ajzen used the term "attitude" to refer to it. Thus, an attitude toward an "object" is defined as the positive or negative affect that one feels toward that object. The third set of constructs used to represent attitudes in the literature were people's intentions to behave in certain ways with respect to the attitude object. For example, some people intend to vote for Republicans, intend to donate money to Republicans, intend to campaign against Republicans, and so on. Fishbein and

Ajzen referred to these constructs as "behavioral intentions." Fishbein and Ajzen offered an insightful analysis showing that the kinds of factors that influence beliefs are different from the kinds of factors that influence attitudes, which, in turn, are different from the kinds of factors that influence behavioral intentions. Hence, they argued, it makes scientific sense to distinguish between beliefs, attitudes, and behavioral intentions instead of treating them interchangeably as reflecting the same construct.

As another example, there is controversy in sociology and anthropology about ways to define and conceptualize social class. Qualities or characteristics used in the many definitions of social class include occupation, education, income, wealth, land/property ownership, control over means of production, political standing, prestige, reputation, associations with elites, honorary titles, and language, among others. Stratum approaches to social class group people into different classes, including two-class models (powerful and weak), three-class models (lower, middle, and upper class), and multistratum models having as many as nine classifications (Fussell, 1983). The classic work of Karl Marx defines class in terms of the degree of control over the means of production (Marx, 1887). Marx also emphasized subjective class (class identity) versus objective class (actual relationship to the means of production). The sociologist Max Weber (1904) defined social class in terms of three components: economic considerations (relative to control over means of production), prestige, and political position. The various characterizations of social class share the assumption that society is stratified in terms of economic-related variables (representing shared meaning), but the many characterizations of social class tend to use additional attributes as a basis for classification, depending on the theory (representing surplus meaning).

Which conceptualization of social class is optimal? The answer to this question is best viewed from the perspective of Meehl's classic work on nomological networks (Cronbach & Meehl, 1955; MacCorquodale & Meehl, 1948). According to this framework, the meaning and utility of a concept emerges in the context of the broader theoretical network in which the concept is embedded. Thus, the "most appropriate" definition of social class depends on how the construct is used in the theory being advanced and how it fosters coherence in the overall theory. Of course, it is preferable if conceptual definitions of a construct across different theoretical frameworks have shared as opposed to surplus meaning. But the meaning and worth of a construct ultimately depends on the broader nomological network in which it is embedded. This viewpoint is consistent with recent views of distributed meaning and situated cognition as ways of determining the meaning of constructs (e.g., Hannerz, 1993; Hutchins, 1996).

PRACTICAL STRATEGIES FOR SPECIFYING CONCEPTUAL DEFINITIONS

A reasonable first step for developing a conceptual definition of a construct is to examine the scientific literature to see how other scientists have done so. A comprehensive and careful literature review of previous research that has used the concept will often yield one or more conceptual definitions that other scientists have found to be useful.

Often, the difficult task of specifying an unambiguous definition will have already been performed. This is not to say that you should, by fiat, accept the conceptual definitions offered in previous work. Perhaps, from your perspective, the definitions miss the essence of the concept or are flawed in one manner or another. However, reflecting on the definitions of others who have thought long and hard about the concept will often be fruitful.

Another useful source is a dictionary. You might be surprised at how often a clear and concise conceptual definition of a term can be found in a dictionary (or that might be suggested by looking up the term in a thesaurus). We make it a standard practice to look up any terms for constructs we plan to use in our conceptual system in two or three major dictionaries. To be sure, scientific definitions sometimes deviate from everyday definitions found in dictionaries, but consulting a dictionary may help you in the process of specifying a definition with better clarity and insight.

Another strategy for devising a clear conceptual definition is to list the key *properties* of a concept. Properties are the identifiable characteristics of concepts and form the cornerstone for later measurement of the concept and empirical testing. Tables possess such properties as height, width, length, type of composition, and color. When describing a table, we do so in terms of these properties. In practice, most concepts can be described by using scores of different properties. The instantiation process serves both a select-in and a select-out function. It tells us what we must focus on when structuring definitions of our concepts, and, at the same time, it puts blinders on us with respect to the remaining aspects of the theoretical expression. Some properties are crucial for conveying the essence of the concept, whereas others may be of less importance. You might try to make a list of the core properties and then assign priorities to them.

As an example, behavioral prediction is central to many theories in the social sciences. Many behaviors have four core elements or properties: (1) an action (e.g., talking about drugs), (2) an object or target toward which the action is directed (e.g., to your teenage daughter), (3) a setting (e.g., in your home, while sitting at the kitchen table), and (4) a time (e.g., on Monday night). When researchers measure or conceptualize a behavior, they implicitly, if not explicitly, commit to treating these behavioral elements at some level of specificity or abstraction. For example, the behavior "the number of alcoholic drinks a person has consumed in the past 30 days" ignores or collapses the settings in which the drinking occurs as well as specific times at which drinking occurs. A feature of time is invoked in this example by requiring that the drinking occur in the past 30 days. In addition, the object (alcoholic drinks) represents an abstract category that subsumes multiple instantiations of that category (e.g., drinking beer, wine, hard liquor). Research has affirmed the theoretical importance of being explicit about how each of the four elements of a behavior is treated because the relevant predictors and determinants of that behavior can vary depending on the level of abstraction of the elements (e.g., Ajzen & Fishbein, 1980; Jaccard, 1974). This, in turn, affects the nature of the theory you develop. Thus, providing a clear definition of an outcome behavior of interest involves addressing the four properties of action, target object, setting, and time.

Another heuristic for delimiting the nature of concepts is to successively ask the question “What do you mean by that?” For example, for the concept of “attitude toward African Americans,” one might answer the question “What do you mean by that?” as follows: “It is a tendency to act in a consistently favorable or unfavorable manner toward African Americans.” One would then focus on key words within the answer, asking the same question: “What do you mean by that?” For example, we might ask, “What do you mean by ‘consistently’?” or “What do you mean by ‘African Americans’?” If in response to the latter question, the answer is “a person whose ancestry is predominantly Negroid,” you might further ask “What do you mean by ‘predominantly’?” and “what do you mean by ‘Negroid’?” And so on. This process could continue forever, eventually leading to a linguistic nightmare. The idea is to continue with this line of reasoning until you judge that all major ambiguities are clarified and that the meaning of the construct can be clearly communicated.

Another strategy for specifying a concise conceptual definition is to play the role of a journalist who, in the context of an article for a magazine or newspaper, must explain the nature of the variable and its meaning to the public. Write the definition as if you were writing the article. Using this strategy can help you avoid jargon that may carry with it surplus meaning. Even when a conceptual definition is well understood by the scientific community, engaging in this exercise may prove to be useful to ensure that you can articulate all aspects of the definition in a meaningful fashion.

Yet another strategy that can be used to avoid ambiguous jargon is to place yourself in the position of having to define and explain the concept to someone who is just learning the English language, thereby forcing yourself to use more simplified and commonplace terms.

Using a denotive definition strategy is another strategy for articulating a definition (McGuire, 1997). For example, if you want to provide a definition of “political conservatism,” you might assign familiar political figures to different locations on the conservatism dimension and then try to specify which characteristics of each politician made you place him or her where you did.

Creating precise conceptual definitions can also be achieved by articulating or writing out how the concept would be measured in an empirical investigation. Ambiguities in the construct often reveal themselves when the construct must be operationalized at such a concrete level. For example, if you are trying to define a personality trait such as dominance, how would you go about assessing or measuring it in people? Exactly what questions would you ask? As you formulate these questions, the meaning of the concept often will become clearer to you.

A final strategy for specifying a clear conceptual definition is to use the principles of grounded theory construction. We defer consideration of this strategy until Chapter 10.

A common trap when offering conceptual definitions is to not provide a definition but, rather, to define the concept by providing examples of it. In general, this tactic should be avoided. To be sure, specifying exemplars can be a useful strategy to help you think through the key properties and characteristics of a concept and to help convey a sense of the concept to others. However, such exemplars should not be a substitute for a carefully worded and clear conceptual definition.

MULTIDIMENSIONAL CONSTRUCTS

When defining concepts in a theory, it sometimes is useful to think of subdimensions or “subtypes” of the construct. Consider the case of intelligence, a widely used concept in the social sciences. Psychologists became actively interested in the scientific study of intelligence in the early 1900s. At that time, intelligence was conceptualized as a rather broad concept, usually in terms of the capacity to solve complex and abstract problems. Such a conceptualization quickly came under attack as being too abstract and theoretically limiting. Several lines of evidence were used to argue against the global conceptualization. The first was a phenomenon known as “savant syndrome”—that is, significantly developmentally disabled individuals with one or more highly developed skills. The most intensive study of this phenomenon, by Scheerer, Rothmann, and Goldstein (1945), documented the case of an 11-year-old boy who was healthy and showed no signs of neurological disturbance. He was unable to perform well in school, his general information was quite substandard, he knew the meaning of few words, and he showed virtually no abstract problem-solving skills. His IQ on the Stanford–Binet (a widely used intelligence test) was 50, a very low score. However, there also were some paradoxes. The boy could tell the day of the week for any given date between 1880 and 1950. He could correctly spell many words forward and backward, and he never forgot the spelling of a word once he was told how to spell it. He could play by ear many complex musical compositions and could sing the opera *Othello* from beginning to end. The savant syndrome phenomenon suggested the need to recognize more specialized capacities in the context of the concept of intelligence.

A second perspective on the inadequacy of a general conceptualization of intelligence was provided by a series of factor-analytic studies. These studies postulated different types of intelligence and demonstrated empirically the utility of separating intelligence into different forms. One of the more popular conceptualizations is that of Leon Thurstone (1947), who postulates seven primary mental abilities: number, word fluency, verbal meaning, memory, reasoning, spatial perception, and perceptual speed. The results of factor-analytic studies underscore the idea that general concepts frequently need to be made more specific, and that it is possible to use empirical strategies to help scientists reconceptualize their constructs.

Examples of multidimensional conceptualizations of constructs in the social sciences abound. For instance, theories of risk taking have delineated four types of risk-taking propensities of individuals: (1) physical risk taking (putting oneself in harm’s way, physically), (2) social risk taking (taking risks in social relationships), (3) monetary risk taking (taking risks with money), and (4) moral risk taking (taking risks involving the breaking of rules or laws). These “components” of risk taking derive from the proposition that risk taking can occur in different settings and that risk tendencies can vary

BOX 5.1. Etic and Emic Constructs

In anthropology distinctions have been made between etic- and emic-based approaches to scientific analysis. The terms *emic* and *etic* were coined by the linguist Kenneth Pike and were popularized in anthropology by Marvin Harris. The terms are applied in many different contexts, with somewhat different meanings depending on the subdiscipline. The essence of emic approaches is to understand a culture in the way that members of that culture understand it, to learn the concepts they use, and to try to see the world as they do. By contrast, the essence of etic approaches is to understand a culture in more abstract scientific terms that apply across different cultures and that can be used to make cross-cultural comparisons. The concepts and theories in an etic approach derive from a comparative framework and can be meaningless to the members of the culture under study.

An emic construct is one that reflects the perceptions and understandings of the members of a culture. For example, the concept of "motherhood" would be defined emically by how members of a cultural group view that role. Etic constructs, by contrast, are defined by scientists in more abstract terms and apply across groups (Headland, Pike, & Harris, 1990). The focus is not on how a specific culture defines "motherhood" but rather on a conceptual definition of the concept of "motherhood" as used by the scientific community.

In social science research, a common orientation is to combine etic and emic approaches to theory and measurement. The idea is to build theories of the relationships between variables at an etic level, but to operationalize and measure those constructs at an emic level. For example, a scientist might be interested in the construct of "expectancies" and define them as the perceived advantages and disadvantages of performing a behavior. The actual content of expectancies can vary from one group to another. For example, adolescent males might define "dating" somewhat differently than adolescent females and they might perceive different advantages and disadvantages of dating than females. The idea is to measure the constructs of dating and expectancies in such a way that these different perceptions and conceptions are respected, but to still apply the etic concept of expectancies to try to predict and understand the etic construct of dating behavior.

Examples abound of emic manifestations of etic concepts. For example, in some cultures it is appropriate to mourn individuals who have recently died by wearing black clothes, whereas in other cultures, one wears white clothes while mourning. When greeting one another, it is appropriate for two men to kiss on the cheek in some cultures, whereas such actions would be appalling in other cultures, where a greeting is confined to a simple handshake.

as a function of setting. Another example: Theories of social support distinguish four types of support people receive: (1) tangible support (e.g., monetary loans, transportation to and from different places, help with child care), (2) emotional support (providing empathy and understanding for one's feelings and emotions), (3) informational support (providing technical or practical information that a person needs to deal with a problem), and (4) companionship support (providing company for doing fun activities). How people cope with problems and stress is thought to be influenced by the four different types of support.

You can often make a theoretical network richer and concepts more precise and clearer by specifying subcomponents or dimensions of a higher-order construct.

CREATING CONSTRUCTS

Social scientists sometimes invent or create variables for the purposes of building new theories. For example, in the 1960s and 1970s, a personality variable was created and popularized by Friedman and Rosenman (1974), called the "Type A" and "Type B" personality syndromes. These syndromes were thought to be risk factors for coronary heart disease. Type A individuals were described as impatient, time conscious, competitive, and as having difficulty relaxing; they are workaholics who are deadline oriented and who internalize stress. Type B individuals were patient, relaxed, and easygoing. There also was a Type AB profile, consisting of people who cannot be clearly categorized into the two types. The Type A versus Type B personality syndromes became the subject of considerable research in the health domain (Friedman, 1996).

Another example is the concept of emotional intelligence (Goleman, 1995; Salovey & Mayer, 1990). As noted earlier, considerable research has been conducted on the concept of intelligence and its ability to predict job and school performance. Intelligence tests are widely used to make placement decisions in schools and to make hiring decisions in organizations. Salovey and Mayer (1990) argued that traditional intelligence tests measure cognitive and problem-solving abilities, but that an equally important ability for success is "emotional intelligence." Emotional intelligence is the ability to monitor one's own and others' feelings and emotions, to discriminate among them, and to use this information to guide one's thinking and actions (Salovey & Grewal, 2005). The core features of emotional intelligence have been subject to hundreds of studies in the social sciences for decades. Salovey and Mayer grouped together some core areas of emotion research, "repackaged" them into a coherent whole, and gave them a provocative label. The result was a noteworthy advance in theoretical and empirical perspectives on emotions (Salovey & Grewal, 2005).

Social scientists use a variety of strategies for "creating" variables. One strategy is to translate an individual-level variable into a contextual-level variable. Using ethnicity as an example, it is possible to characterize ethnicity (a characteristic of an individual) at higher contextual levels, such as the ethnic composition of a school that individuals attend, the ethnic composition of the neighborhood in which the schools are located, the ethnic composition of the city in which the neighborhoods are located, and the ethnic composition of the state within which the cities are located. You then can examine how these multiple levels of context, characterized in terms of ethnicity, impact the behavior of individuals or some other outcome. For example, how are voting preferences of an individual influenced by the political composition of the neighborhood in which the person resides? As another example, you could assess a person's attitude toward smoking cigarettes, the average attitude toward smoking cigarettes of students in the school that a given student attends, the average attitude toward smoking cigarettes of people in the neighborhood where the schools are located, the average attitude toward smoking cigarettes of people in the city where the neighborhoods are located, and the average attitude toward smoking of people in the state where the cities are located. You can then examine how these multiple levels of context impact the behavior of individuals. In short, you can describe a context in terms of an aggregation of any individual-level variable.

Another strategy sometimes used for creating variables is to reframe environmental or contextual variables to represent perceptions on the part of the individual. Thus, rather than studying the characteristics of the family environment within which an individual resides, you might study how individuals *perceive* the family environment. Or, instead of studying the characteristics of the organizational climate of a business, you might study how the individual *perceives* the organizational climate. Or, instead of working with someone's actual social class, you might instead focus a theory on people's perceived social class. Theories contrasting "perceived" versus "actual" states of affairs are a rich source of scientific theorizing. As an example, Dittus and Jaccard (2000) studied the effects of mothers' disapproval of their child engaging in sexual intercourse on whether the child engaged in sexual intercourse in the next 12 months. Two variables were measured: the mother's actual disapproval of the adolescent engaging in sexual intercourse (as measured on the mother) and the child's perception of the mother's disapproval (as measured on the child). Jaccard and Dittus found only a modest correlation between perceived and actual maternal disapproval, suggesting that mothers were not doing a satisfactory job in conveying their expectations to their children. They also found that both variables were independently predictive of future sexual behavior, though the adolescent perceptions were a somewhat stronger predictor than the actual maternal attitudes.

As you build a theory, you might consider creating new constructs using the strategies described above.

AN EXAMPLE OF SPECIFYING CONCEPTUAL DEFINITIONS

It may be useful to consider an example of defining variables that uses some of the principles discussed in this chapter. Suppose we posit a simple theoretical relationship between gender and adolescent alcohol use. We might assert that there are gender differences in adolescent alcohol use, with males tending to drink more than females. The first step in our analysis is to consider in more depth the two variables we have specified and to clarify what we mean by each. One construct is gender (male vs. female). We might decide that the meaning of this variable is so obvious to the scientific community that we do not need to provide a formal definition of it. The meaning of some variables is consensually accepted by all, and it is not necessary to define every term you use. Note, however, that what is consensual in some research areas will not be consensual in other research areas. For example, the concept of gender is a highly controversial one for the Olympic Games. In this context, determining if someone is a male or a female involves complicated issues about hormone levels and basic physiology. Several years ago, a struggling male professional tennis player underwent a sex-change operation and then joined the women's professional tennis tour, to the loud objections of many that he was not really a female. For the purposes of our theory, however, we might assume that gender is a concept whose meaning has social consensus.

The variable of "adolescent alcohol use" is another matter. This is a somewhat abstract and fuzzy concept and we need to be more explicit about what we mean by it. We first need to be explicit about what we mean by *adolescent* and to which adolescents we are referring. We might decide that we want to restrict our statements to early adolescents (middle-school-age youths) living in the United States. In terms of the behavior, we start by using the four behavioral elements identified earlier (action, target, setting, time) and analyze the variable in terms of each element. What is the precise action on which we want to focus? Do we want to focus on acquiring alcohol, consuming alcohol, or both? For the adolescent population about whom we are theorizing, it is illegal to purchase alcohol, so gaining access to alcohol is not a trivial matter. We might decide to focus on the act of consumption and ignore issues of acquisition since, in the final analysis, it is consumption where our interest truly lies. In terms of the target of the behavior, the action is performed with respect to alcoholic beverages. Do we mean a specific type of alcoholic beverage—such as beer, wine, hard liquor, mixed drinks—or do we want to generalize across all forms of alcoholic beverages? We might decide to theorize about the latter. In terms of the setting, do we care about the particular setting in which alcohol consumption occurs (e.g., drinking alone or drinking with other people) or do we want to generalize or aggregate across settings? We might decide to pursue the latter. In terms of a time frame, over what period of time do we want to examine alcohol consumption? Should it be over the past day, during the past week, during the past month, during the past year, or during one's lifetime? Suppose we decide to focus on "current" behavioral tendencies with respect to alcohol consumption and decide that alcohol use during the past 30 days provides us with a good sense of a person's current drinking habits and orientations. The decision to focus on a 30-day interval is not an arbitrary one, but instead is based on the existing literature on alcohol use, which we do not elaborate upon here.

As we think more about quantifying the alcohol variable, we realize there are actually two facets of it we can examine. First, people can differ in terms of the number of drinking episodes they have had in the last 30 days. Second, they can differ in how much alcohol they consume during a given episode. This might lead us to distinguish between three dimensions of alcohol use: (1) the frequency of drinking alcohol, (2) the amount of alcohol consumed at each drinking episode, and (3) the total amount of alcohol consumed (obtained by weighting the number of episodes by the quantity consumed at each episode). We do not consider here the methodological challenges that confront researchers in measuring these different dimensions. We only note that they are theoretically distinct constructs that represent somewhat different dimensions of alcohol use.

There is yet another dimension of alcohol use we might consider and that is getting drunk. Some people consume alcohol but never get drunk because they drink in moderation. Others get drunk and, indeed, go out drinking with the explicit intention of getting drunk. The latter scenario is often referred to as “binge drinking” in the alcohol literature. So, instead of simply theorizing about drinking frequency and the quantity of drinks per episode, we might also decide to focus on the number of times that an individual became drunk in the past 30 days.

Sometimes, even though they may get drunk, people drink without incident. At other times, people drink too much and get into trouble as a consequence. A person might engage in unprotected sex because he or she is drunk. A person might destroy property or commit vandalism while drunk. A student might skip a class the next day because of a hangover and miss an exam. Yet another possible dimension on which we could focus is that of the experience of negative problems as a result of drinking.

It is possible that gender differences exist with respect to all of these dimensions, or there may be gender differences on just some of them. Do males tend to drink more often than females? Do males tend to drink more on a given drinking episode than females? Across all episodes of drinking, do males drink more alcohol than females? Do males get drunk more often than females? Are males more likely to get into trouble because of their drinking than females? Our fairly simple theory about gender differences in alcohol use has become more elaborated as we refine and critically think about the construct of “alcohol use.” Note that we are not obligated to focus our theorizing on all these dimensions. We may choose to focus on all of them or only a subset of them. But it helps in the theory construction process to think through our constructs and then to consciously make choices about where to focus our efforts.

OPERATIONISM

We close this chapter with an important historical note concerning a psychometric controversy that existed in certain areas of the social sciences and that has implications for how we define concepts. Conceptual definitions specify *what* needs to be assessed in empirical science, but the matter of *how* they will be assessed is a distinct issue. This latter function is served by what scientists refer to as *operational definitions*, which are central to the design of empirical tests of a theory. The process of developing an operational definition begins with the consideration of the conceptual definition to which it is addressed. The investigator then devotes serious thought to the kinds of operations and procedures that might be employed to provide a satisfactory indication of the concept in question. Consider the notion of “hunger.” Conceptually it can be defined as a craving or need for food or a specific nutrient. In operationalizing “hunger,” psychologists have (1) asked people to respond to questionnaire items regarding their degree of perceived hunger; (2) deprived individuals of food for differing amounts of time so as to create more hunger in some than in others (e.g., people deprived for 24 hours must surely be hungrier than those deprived for only 2 hours); (3) measured the amount of food consumed from a standard portion given to each study participant (e.g., 2 pounds of spaghetti), under the assumption that the more a participant consumes (adjusted by body weight and metabolism), the hungrier he or she is; and (4) measured the amount of adversity the person will go through to obtain food. All of the above seem to be reasonable procedures for measuring, that is, “operationalizing,” hunger.

Unlike conceptual definitions, which often prove difficult to pin down, operational definitions are more concrete and thereby suggest a greater degree of precision and rigor. For this reason, there was a period when many behavioral scientists felt we should abandon conceptual definitions and restrict science to observable operations. This approach was called *operationism*. As examples, some theorists argued that “intelligence tests were what measured intelligence, and intelligence was that which was being measured by intelligence tests” (Bohrnstedt, 1970, p. 84). As another example, instead of considering six purchases of Brand X in a row to be an *indicator* of brand loyalty in consumer psychology, investigators in marketing considered this *to be* brand loyalty itself. In this way, the concept has no identity separate from the instruments or procedures being used to measure it (Jacoby & Chestnut, 1978).

One frequently cited problem with operationism is that if the concept being measured is synonymous with the measurement procedures being used, then even minute changes in method would produce a new concept. The result would be a proliferation of definitions that would lead to confusion and an inability to communicate effectively regarding the concept. “When this occurs, generalizations involving the construct are impossible to make since there is really no single construct under investigation but, instead, a multitude of constructs” (Bohrnstedt, 1970, pp. 94–95). Strict adherence to operationalism “means that both the results of an [investigation] and the conclusions the investigators derived from it can never transcend the methodology employed” (Chaplin & Krawiec, 1960, p. 4). Thus, our ability to generalize beyond any investigation is completely abridged. As a consequence of such problems, the physical sciences long ago discarded the approach of defining phenomena strictly in terms of their operations, and the behavioral sciences have since followed suit.

SUMMARY AND CONCLUDING COMMENTS

The process of instantiation involves delimiting and more narrowly defining the concepts developed during theorization. This process of clarifying concepts removes ambiguities and clarifies what is meant by the construct in question. A major weakness occurs in a theory if it relies on concepts that are so abstract that it is hard to know what the theorist truly means by them (e.g., a statement such as "personality influences behavior"). The hallmark of a good theory is one in which the concepts are clearly defined and where the theorist provides clear-cut examples of the constructs as he or she moves toward more precise instantiations of them.

Sometimes the same concept is defined differently in different theories. The extent to which conceptual definitions overlap is referred to as the shared meaning of a construct; those portions of conceptual definitions that do not overlap across theories are referred to as surplus meaning. Although it is desirable for constructs to have shared versus surplus meaning, the worth and meaning of a construct ultimately are judged relative to the broader nomological network in which the construct is embedded.

Many strategies can be used to make abstract concepts more precise and to clarify the meaning of your concepts. These include consulting past literature, using a dictionary, specifying key properties, continually posing and answering the question "What do you mean by that?", placing yourself in the role of a journalist to explain the nature of the variable to the public, placing yourself in the position of having to define and explain the concept to someone who is just learning the English language, and thinking about exactly how the construct in question would be measured in a scientific study. Another strategy is to define subcategories or subdimensions of a construct and then to clearly define and articulate those. It is not enough to simply provide an example of your concepts. Rather, you must clearly define them.

The present chapter has provided perspectives on how to focus and develop conceptual definitions. The next chapter addresses how to focus and develop relationships between constructs using the tool of thought experiments.

SUGGESTED READINGS

Headland, T., Pike, K., & Harris, M. (1990). *Emics and etics: The insider/outsider debate*. Newbury Park: Sage.–A debate about, and discussion of, etic–emic approaches.

MacCorquodale, K., & Meehl, P. (1948). On a distinction between hypothetical constructs and intervening variables. *Psychological Review, 55*, 95–107.–A classic article on types of variables.

Nunnally, J., & Bernstein, I. (1994). *Psychometric theory*. New York: McGraw-Hill.—A classic on psychometrics, including a discussion of the importance of conceptual definitions.

Theory and Psychology, February 2001, Volume 11.—A special issue containing numerous articles debating operationism.

Wicker, J. (2004). A theory of formal conceptual definitions: Developing theory-building measurement instruments. *Journal of Operations Management*, digital version. A discussion of the need for formal conceptual definitions and how to develop better measurement instruments for theory building through their specification. Available at *www.elsevier.com/wps/find/journaldescription.cws_home/523929/description*

KEY TERMS

instantiation (p. 75)
hypothesis (p. 76)
conceptual definition (p. 77)
focal problem statement (p. 77)
nomological network (p. 78)
shared meaning (p. 78)
surplus meaning (p. 78)
properties of a concept (p. 80)
multidimensional concepts (p. 82)
operational definition (p. 87)
operationism (p. 88)

EXERCISES

Exercises to Reinforce Concepts

1. What are the major functions of instantiation—that is, why do scientists use it?
2. How is instantiation similar to stating a hypothesis or a focal problem statement?
3. What is the difference between shared meaning and surplus meaning?
4. Describe the major strategies for making a fuzzy concept less fuzzy.
5. What is the difference between an operational definition and a conceptual definition?
6. What are the primary objections to operationism?

Exercises to Apply Concepts

1. Choose a concept, provide a conceptual definition of it, and then describe the process you used to create the conceptual definition.
2. Find an example in the research literature of a construct with a fuzzy conceptual definition and provide a clearer definition.

6

Clarifying Relationships Using Thought Experiments

It is sometimes important for science to know how to forget the things she is surest of.

—JEAN ROSTAND (1958)

Theories often involve relationships between concepts. For example, one might posit that there are gender differences in smoking, such that males tend to smoke more than females. This implies a relationship between "gender" and "smoking." Or a theory might state that there is a relationship between intelligence and grade-point average in school, such that higher levels of intelligence are associated with higher grade-point averages.

The above examples identify relationships between variables—a common occurrence in theories. Even theoretical statements that do not seem to invoke variables often implicitly do so. For example, consider the statement "Males are tall." This statement involves two concepts, gender and height. When we say "Males are tall," the immediate question becomes "Tall relative to whom?" If you pursue the underlying logic with the theorist who makes the statement, he or she usually will say that males are tall relative to females, thereby formally invoking the variable of gender. And if you ask the person what he or she means by "tall," he or she will invariably respond with a description of the variable of height. As another example, consider the statement "I dress for the weather." This statement expresses a relationship between what the weather is and how a person dresses: When it is cold, the person wears clothing that keeps him or her warm, and when it is hot, the person wears clothing that maximizes coolness.

Just as concepts can be too abstract, fuzzy, or poorly defined, so can relationships. As you construct a theory that specifies relationships between variables or constructs, it is important to be clear about each relationship you posit. You need to think through the relationship carefully and be able to describe it to others, unambiguously. In this chapter we use thought experiments to help you clarify relationships in your theory. These thought experiments are conducted independent of actual data. They are conducted mentally, usually before data are collected, thereby serving as a heuristic device to clarify your a priori logic about relationships between constructs.

We begin by considering the use of thought experiments in grounded and emergent theorizing and the variable-centered nature of the thought experiments as used in this chapter. Next, we make distinctions between different types of variables that are important for conducting the thought experiments. These include the distinction between categorical and quantitative variables. We then describe thought experiments for articulating the relationships between two variables, considering first the case of the relationship between two categorical variables, then the relationship between two quantitative variables, and then the relationship between a categorical and a quantitative variable. Finally, we introduce the concept of moderated relationships and describe thought experiments for clarifying them. As we describe relationships through thought experiments, we often encourage you to think about two features of a relationship, namely, (1) the nature of the relationship and (2) the strength of the relationship. With the introduction of moderated relationships, we address yet a third facet of relationships: the stability or generalizability of the relationship.

THOUGHT EXPERIMENTS FOR RELATIONSHIPS IN GROUNDED AND EMERGENT THEORY

The thought experiment strategies in this chapter emphasize variable-centered approaches to analyzing relationships. Some grounded theorists prefer process-oriented explanations, describing how individuals or entities get from point or event *A* to point or event *B*, such as the process or steps an organization goes through in moving from a small, fledging start-up company to a large, successful corporation. This is a different mindset than describing relationships between variables. Although some social scientists associate grounded and emergent theories with an exclusive focus on process-oriented explanations, the fact is that such theories tend to include both process-oriented *and* variable-oriented explanations. This is a strength, because the advantages of both approaches are exploited. By contrast, predominantly variable-oriented approaches tend to dominate many areas of the social sciences. As discussed in Chapter 4, we believe such approaches could benefit from adopting process-oriented perspectives as well.

The blending of the approaches is evident, for example, in analyses of expert decision making by government officials overseeing enforcement of environmental regulations. A process-oriented approach might describe the decision process as moving from the initial stage of problem recognition, where the official realizes that a specific environmental problem needs to be addressed, to the final stage of solution implementation. In between, decision makers move through a series of steps, phases, or activities, such as identifying their goals, defining the possible solutions to the problem, gathering information about the possible solutions, evaluating the different solutions in terms of their strengths and weaknesses, making a choice as to which solution to implement, and then implementing the chosen solution. Some officials proceed through these activities in the sequence just described, whereas others intermingle them in complex ways. A process-oriented theory might pursue detailed analyses of the activity patterns of various officials when making decisions and factors that impact on movement from one activity to another.

Variable-centered orientations come into play as theorists realize that at a given stage of the process, there are variations in how officials behave. For example, some officials recognize a problem that needs attention and immediately jump into action to address it, whereas other managers realize that there is a problem to address but they procrastinate before addressing it. The theorist wonders why this is the case. What makes some officials procrastinate but other officials not? Perhaps it is because some officials are preoccupied with other problems. This latter question and plausible answers to it fall in the realm of variable-oriented explanations, but it is every bit as interesting to an emergent theorist as the broader process analysis that is organizing the research.

Whether theorists identify their approaches as emergent or confirmatory, use methods that are qualitative or quantitative, write their theoretical prose in narrative or mathematical form, or identify their primary orientation to explanation as process- or variable-centered, the fact is that most theories inevitably engage in some degree of variable-oriented analysis of relationships. As such, the thought experiments described in this chapter can be helpful to all theorists as they seek to clarify in their minds relationships between variables. We return to process-oriented approaches in future chapters, especially in Chapter 10.

The thought experiments can be conducted at any time during the theory construction process. For example, they could be conducted before data are collected to clarify your a priori thoughts, or they could be conducted as you are analyzing data, to clarify your thinking about what is tentatively emerging. Sometimes relationships will be so straightforward that you will not need to apply a thought experiment. Indeed, some of the examples in this chapter are of this character, but we chose them to better illustrate the logic of structuring a thought experiment.

DESCRIBING RELATIONSHIPS WITH DIFFERENT TYPES OF VARIABLES

The way a theorist characterizes a relationship between two variables in a thought experiment differs depending on the type of variables involved. An important distinction is whether a variable is categorical or quantitative. A categorical variable has different "levels," "values," or "categories," but there is no special ordering to the categories along an underlying dimension. For example, gender is a categorical variable that has two values or levels, "male" and "female." Religious affiliation is a categorical variable that has the categories "Catholic," "Protestant," "Jewish," "Muslim," and so on. The categories are merely labels that differentiate one group from another. Other terms used to refer to a categorical variable are *qualitative variable* and *nominal variable*.

In contrast, a quantitative variable in social science research is one in which individuals are assigned numerical values to place them into different categories, and the numerical values have meaning in that they imply more or less of an underlying dimension that is of theoretical interest. For example, married couples can be classified into the number of children they have, with some couples having a score of 0, some a score of 1, some a score of 2, and so on. There are different types of quantitative variables (e.g., discrete vs. continuous) and measures of them (e.g., ordinal vs. interval), but the distinctions between them are not critical for our purposes.

It is not necessary to adopt causal thinking to describe relationships, but it is easier for us to explain the heuristics if we do. For this reason, this chapter is built around causal theories, that is, conceptualizations where one variable in the theory, *X*, is thought to influence another variable, *Y*.

THOUGHT EXPERIMENTS FOR RELATIONSHIPS BETWEEN CATEGORICAL VARIABLES

Categorical Variables with Two Levels

Consider gender and political party identification, where gender has the values "male" and "female" and political party identification has the values "Democrat" and "Republican." Suppose that as a theorist, you want to argue that gender impacts political party identification such that females are more likely than males to be Democrats and males are more likely than females to be Republicans. We can translate this presumed relationship into a thought experiment to help us articulate it in more depth. The thought experiment requires that you construct a two-way table of frequencies, also called a *contingency table*. Conceptualize a hypothetical group of 100 females and 100 males and create a hypothetical contingency table for the two variables that looks like this:

	Democrat	Republican	Total
Females			100
Males			100

The "cause" (in this example, gender) is written as rows and the "effect" (political party identification) is written as columns. In this case, a person's gender is assumed to influence the political party with which he or she identifies. Next, fill in the blank cells of the table with frequencies that you think reflect the relationship you posit. That is, make up numbers and write them in the table so that they conform to your theoretical proposition. To illustrate, suppose we filled in the table as follows:

	Democrat	Republican	Total
Females	70	30	100
Males	30	70	100

Look at the row for females. It shows that 70 of the 100 females are Democrats and 30 of them are Republicans. Note that this is the same as saying that 70% of the females are Democrats and 30% of the females are Republican because the total number of females equals 100; 70 out of 100 is 70% and 30 out of 100 is 30%. Now examine the row for the males. Here 30% of the males are Democrats and 70% are Republicans. The table reveals the trend put forth in the theoretical statement: Females are more likely than males to be Democrats (because 70% of females are Democrats, but only 30% of the males are), and males are more likely than females to be Republicans (because 70% of the males are Republican, but only 30% of the females are). The hypothetical contingency table in this thought experiment makes the theoretical statement concrete.

Note that, instead of the 70% and 30% figures, we could have captured the same dynamic with a different set of numbers:

	Democrat	Republican	Total
Females	51	49	100
Males	49	51	100

It is still the case in this table that females are more likely than males to be Democrats (51% vs. 49%) and that males are more likely than females to be Republicans (51% vs. 49%). However, now the relationship between gender and political party identification is much weaker than before.

Yet another set of numbers that accurately captures the theoretical statement is this one:

	Democrat	Republican	Total
Females	95	5	100
Males	5	95	100

For this table, the effect of gender on political party identification is extremely strong, because virtually all females are Democrats and virtually no males are Democrats, and vice versa, for Republicans. In terms of what you theorize, does this sound reasonable to you?

When developing a theory that posits a relationship between two categorical variables, we believe it is instructive to pursue thought experiments with hypothetical contingency tables of the form described above and to insert your best guesses as to how you think the percentages would pattern themselves down and across the columns and rows. This will force you to be explicit about your thinking as you try to derive meaningful numbers. It also will give you a sense of how strong you expect your relationship to be. We realize that the numbers that you insert will be somewhat arbitrary and that, at times, you will have little basis for positing a small effect, a medium effect, or a large effect. But we have found it useful at this stage of theory construction to fill out the table and commit to some tentative numbers in order to force ourselves to think through the relationship. As you generate numbers and try to justify them to yourself, it may stimulate you to think of things you might not otherwise have thought about. We emphasize that this is just a heuristic device to be used in the early stages of theory construction as a way of helping to focus your thoughts and make sure you are clear about the relationship you are specifying. When formally describing your theory to other scientists, you typically will not present the hypothetical contingency table that helped clarify your thinking.

To formalize this exercise somewhat, here are the steps we suggest you use when using a thought experiment to clarify the presumed relationship between two categorical variables:

1. Create a contingency table, listing the values of the "cause" as rows and the values of the "effect" as columns.
2. Think of 100 hypothetical people for each row of the table; that is, set the marginal frequency of each row to equal 100.
3. Of these 100 people, specify how many you think will fall into each column category; this represents the percentage of people in each category.

As you complete the numbers in a hypothetical contingency table, always try to verbalize the rationale for your numerical assignments. Doing so will help you defend your theoretical proposition and explicate its underlying logic.

When you embark on the theory construction process, you may have an intuitive sense that two categorical variables are related, but you may not be sure how to articulate that relationship. In such cases, it might help to construct the hypothetical contingency table and fill in the numbers based on your intuition. After doing so, the numbers in the table will make the relationship you have intuitively considered more explicit. With the filled-in contingency table in hand, try to better verbalize the rationale for the relationship and the bases for the numbers you provided. Below we discuss how to derive formal theoretical propositions from a completed hypothetical contingency table. This principle applies to all of the thought experiments described in this chapter.

On the other hand, you may have a well thought-out proposition before you construct the hypothetical contingency table. In this case, you can still use thought experiments to generate numbers, because doing so will (1) force you to think through the proposition in more detailed ways, and (2) force you to think about the strength of the relationship.

Categorical Variables with More Than Two Levels

Sometimes the categorical variables with which you work, either the "cause" or the "effect," have more than two levels. You still approach this situation using the hypothetical contingency table, but now the description of the theory becomes more complicated. As an example, consider the relationship between religious affiliation and political party identification in which each variable has three levels. Here is the relevant table:

	Democrat	Republican	Independent	Total
Catholic				100
Protestant				100
Jewish				100

And here are numerical entries we might generate:

	Democrat	Republican	Independent	Total
Catholic	50	25	25	100
Protestant	25	50	25	100
Jewish	50	25	25	100

From this table, we derive the following theoretical propositions:

Proposition 1: Catholics, Protestants, and Jews are equally likely to be Independents

Proposition 2: Protestants are more likely to be Republicans than either Catholics or Jews; Catholics and Jews are equally likely to be Republicans.

Proposition 3: Catholics and Jews both are more likely than Protestants to be Democrats; Catholics and Jews are equally likely to be Democrats.

The strategy we used for stating the propositions was to focus on each column of the contingency table separately, then state what trends were apparent in the numbers we assigned within that column across the different row categories. For example, Proposition 1 focuses on the column for Independents, Proposition 2 focuses on the column for Republicans, and Proposition 3 focuses on the column for Democrats. Reread the propositions more carefully and take note of how we are making statements about row differences within each column. This is a useful strategy for deriving theoretical propositions from contingency tables.

In sum, a heuristic for clarifying a relationship between two categorical variables is to conduct a thought experiment in which you construct a hypothetical contingency table and then to complete cell frequencies using the proscribed methods. In doing so, articulate the rationale behind the numerical assignments.

THOUGHT EXPERIMENTS FOR RELATIONSHIPS BETWEEN QUANTITATIVE VARIABLES

Instead of two categorical variables, the theoretical relationship you want to clarify might involve two quantitative variables. When this is the case, the thought experiment is pursued using a different graphical device, called a *hypothetical scatterplot*.

Scatterplots

A *scatterplot* is a graph that plots scores or values on one variable, *Y*, as a function of scores or values on the other variable, *X*. In this section we first provide a sense of what a scatterplot is and then we show how to use it in a thought experiment to help clarify a relationship. We also develop the idea of a linear relationship between variables.

Figure 6.1 presents a scatterplot using a simplistic but pedagogically convenient example relating the number of hours individuals work in a week to the amount of money they earn in that week. In this example, there are five individuals, each paid \$5 an hour. Here are the data:

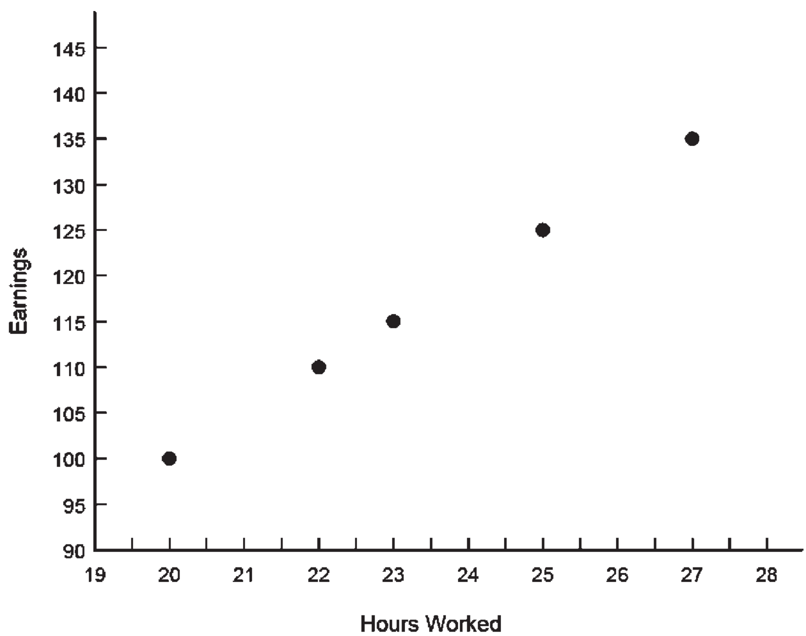


FIGURE 6.1. Scatterplot for earnings as a function of hours worked.

Individual	Hours worked	Amount earned
1	20	$100
2	22	$110
3	25	$125
4	23	$115
5	27	$135

The *X* values (the cause) are always plotted on the horizontal axis and the *Y* values (the effect) are always plotted on the vertical axis. Consider the first individual. We locate that individual's score on *X* (20) on the *X* axis and then we also locate the individual's score on *Y* (100) on the *Y* axis. We draw an imaginary line up from the located point on the *X* axis and across from the located point on the *Y* axis (as illustrated in Figure 6.2). We then draw a dot where the two lines intersect. We repeat this procedure for each individual, until the pairs of scores for all individuals have been plotted with dots. The completed scatterplot is presented in Figure 6.1.

Note that there is a systematic relationship between the two variables: The more hours that someone works, the more money the person earns. If we connect the dots that demarcate each individual's pair of scores, we see a straight line, as in Figure 6.3. This is an example of what is called a *linear* relationship.

It is rare in the social sciences for two variables to have a perfect linear relationship, but sometimes they approximate linear relationships. Figure 6.4 presents a scatterplot showing height and weight for 15 individuals. You can see that the data approximate a

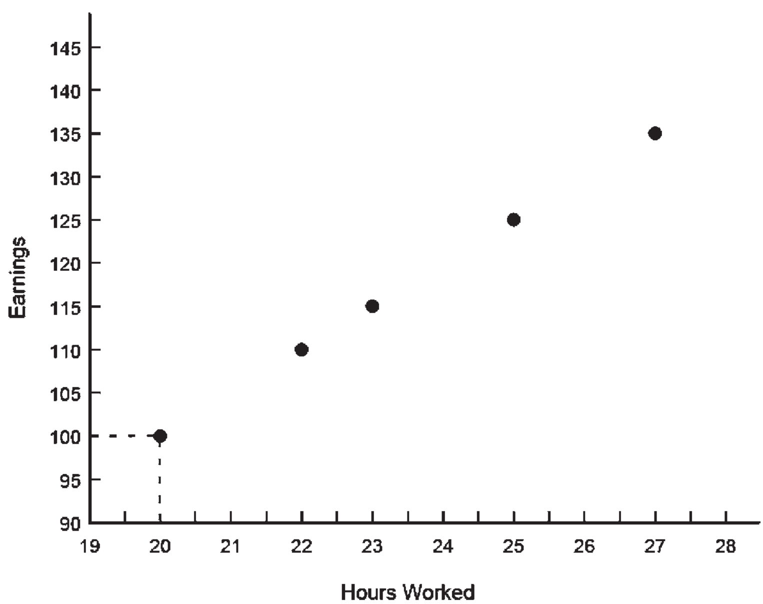


FIGURE 6.2. Scatterplot with dotted lines for earnings as a function of hours worked.

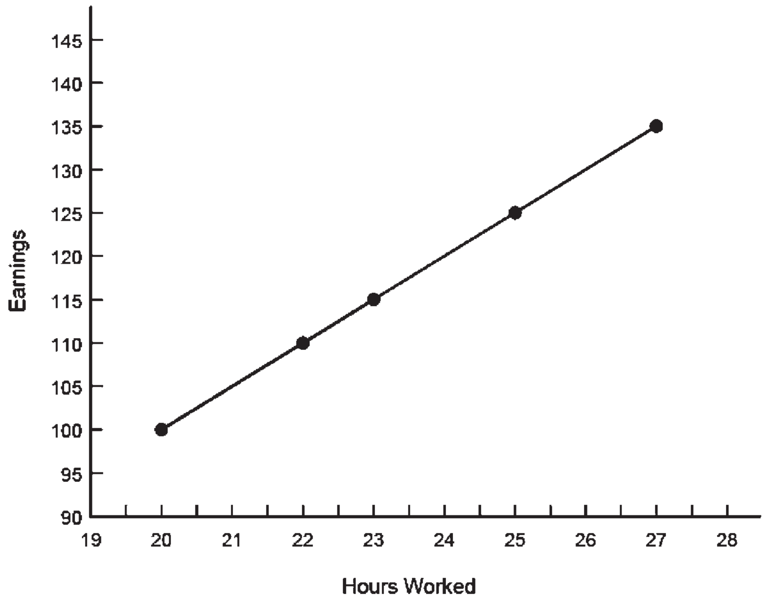


FIGURE 6.3. Scatterplot with dots connected for earnings as a function of fours worked.

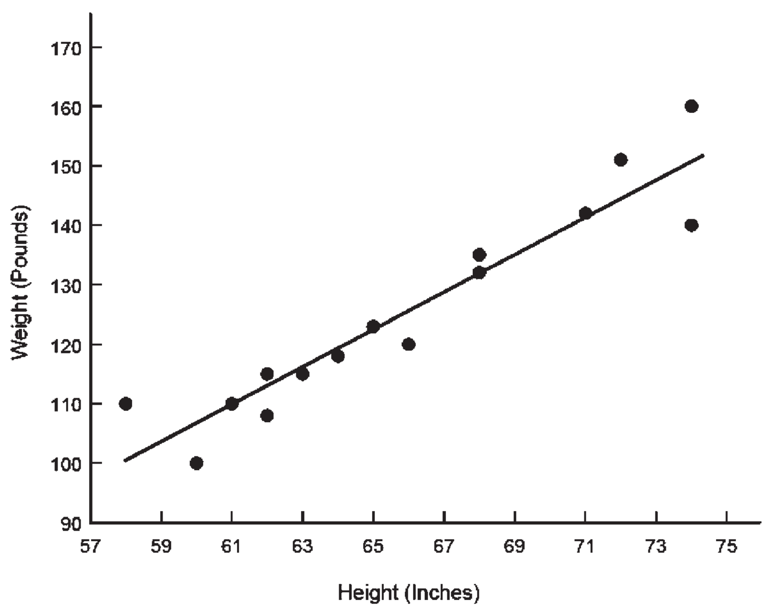


FIGURE 6.4. Scatterplot between height and weight.

linear relationship. We have drawn a line through the data on the scatterplot to emphasize this trend. If there was a perfect linear relationship between the variables, all the points would fall on a straight line. They do not in this case, but there is enough of an approximation that we can reasonably talk about the relationship as if it were linear. The deviations from linearity appear to reflect random "error." (We consider the case of systematic nonlinear relationships later.)

Characteristics of Linear Relationships

Linear relationships can be of three types. First, there is a *direct linear relationship* where higher scores on *X* imply higher scores on *Y*. This is the case in Figures 6.1 and 6.4. Second, there is an *inverse linear relationship*, where higher scores on *X* imply lower scores on *Y*. This case is illustrated in Figure 6.5, which represents the relationship between people's satisfaction with their job and how many days of work they miss (not counting vacations and holidays) over a period of 12 months. The greater the degree of job satisfaction, the fewer the number of days missed. Another example of an inverse relationship is one between amount of exercise people get and the probability that they will have an early heart attack. The more people exercise, the lower the probability that they will have an early heart attack. A *zero-slope linear relationship* is characterized by a flat line. It means that as *X* increases, *Y* is predicted to remain at the same value. Figure 6.6 provides an example of such a relationship. It plots the age of toddlers and how much their mothers say they love them, as rated on a scale from 1 to 5, with higher numbers

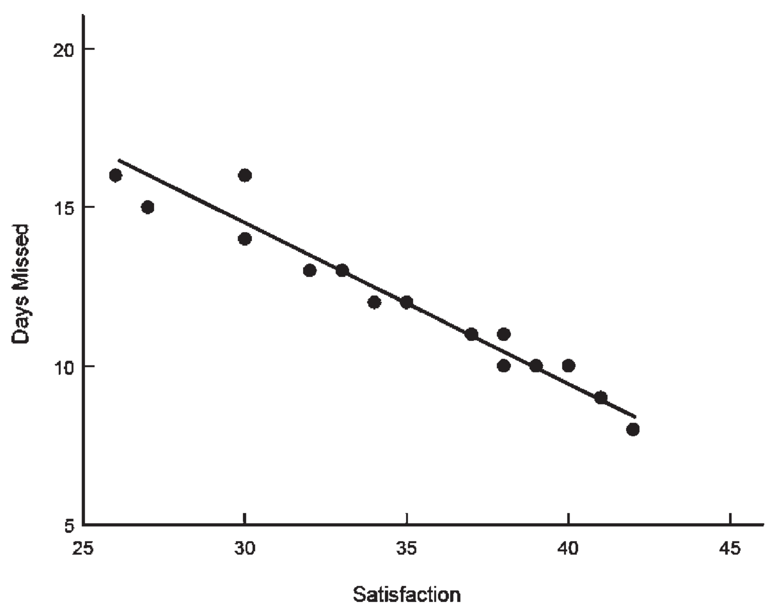


FIGURE 6.5. Example of an inverse linear relationship.

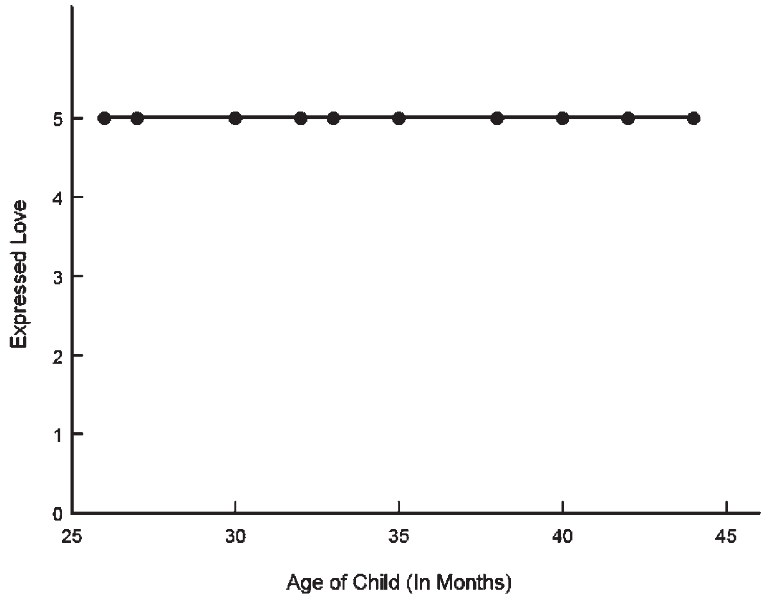


FIGURE 6.6. Example of a zero-slope relationship.

indicating more love. No matter the age of the toddler, the mother's reported love does not differ.

When describing a relationship between two quantitative variables, theorists often do so with reference to a direct linear relationship, an inverse linear relationship, or a zero-slope relationship. Although we recognize that data seldom are perfectly linear in form, a shorthand heuristic for visualizing the relationship between two variables is to draw a line on a scatterplot that reflects either a direct linear relationship, an inverse linear relationship, or a zero-slope linear relationship, as has been done in Figures 6.7a, b, and c. The values of *X* and *Y* are not written on the axes; instead scores are indicated as going from "low" to "high" on each axis (again, this is just a shorthand heuristic device). When drawing lines, the steeper you draw the slope of the line, the larger the effect of *X* on *Y*, everything else being equal. Figure 6.8 shows a case of two lines on the same graph that have different slopes, with line *A* reflecting a stronger effect than line *B* (assuming the scales for the two lines are on a common metric). A change in *X* for line *A* produces a greater change in *Y* than a comparable change in *X* for line *B*. These are examples of *shorthand hypothetical scatterplots*.

Nonlinear Relationships

Not all relationships between quantitative variables are linear; some are nonlinear. Consider the shorthand scatterplot in Figure 6.9. This is an example of a nonlinear relationship, called an *inverted-U-shaped relationship* (because it looks like a U drawn upside down). Some quantitative relationships between variables take this form. For example, levels of anxiety about taking a test and performance on that test are often characterized

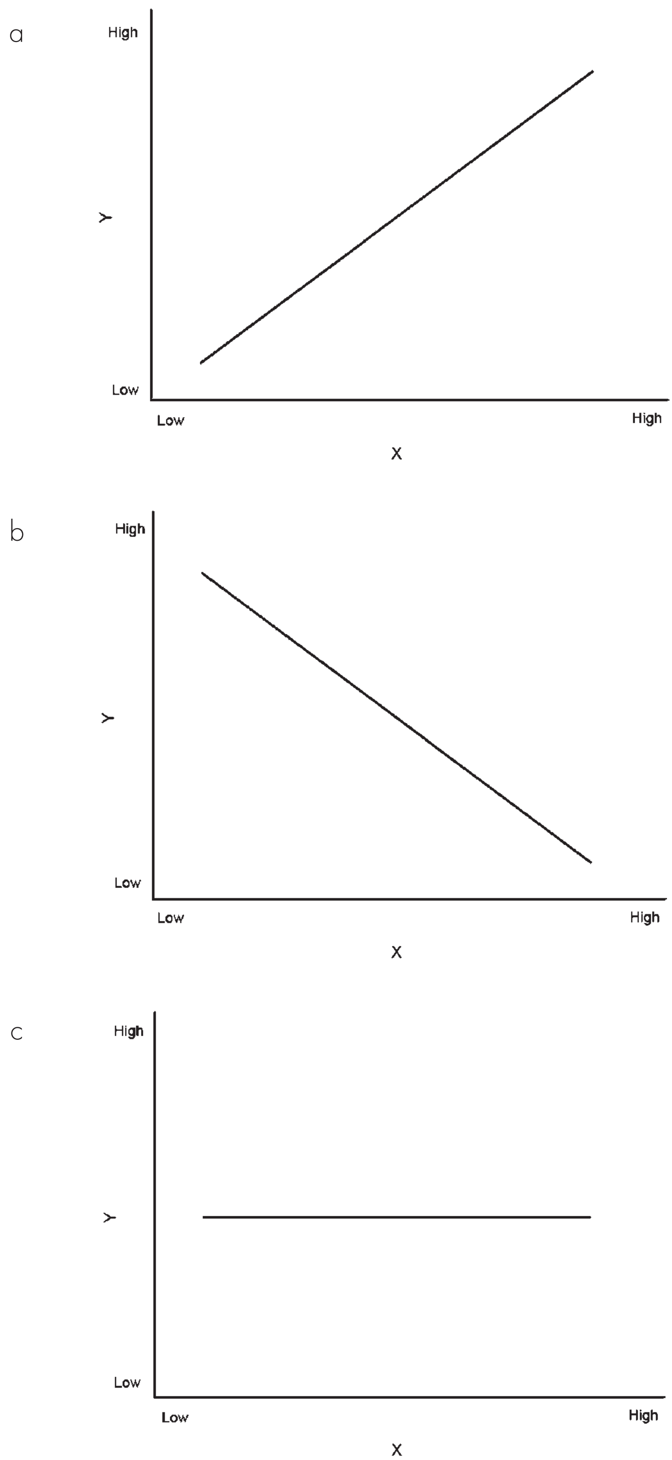


FIGURE 6.7. Examples of hypothetical scatterplots. (a) Direct linear relationship; (b) inverse linear relationship; (c) zero-slope relationship.

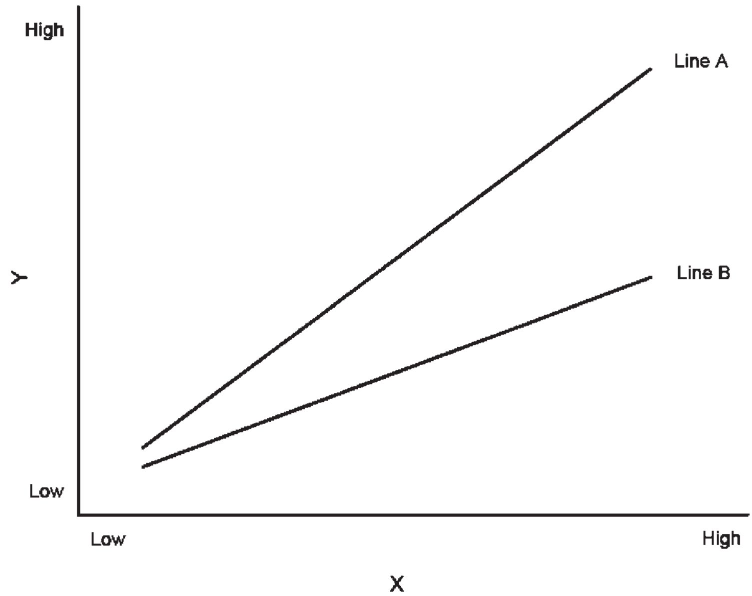


FIGURE 6.8. Example of lines with different slopes.

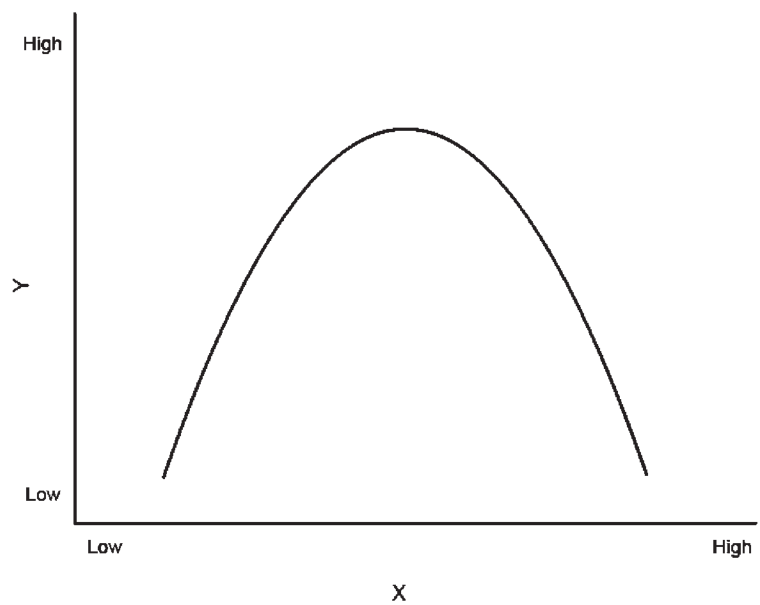


FIGURE 6.9. Example of inverted-U function.

by an inverted-U function. The idea is that increasing anxiety at low levels of anxiety is good and facilitates test performance by increasing the person's attention and arousal. However, at some point, higher levels of anxiety get in the way and start causing the person to worry needlessly and become distracted from the task at hand. This response, in turn, degrades test performance. The result is an inverted-U relationship between test anxiety and test performance.

Figure 6.10 illustrates another common nonlinear relationship, sometimes referred to as an *S-shaped relationship*. At low levels of *X*, there is a *floor effect* such that changes in *X* have no impact on *Y*. Then at some point, increases in *X* begin to lead to increases in *Y*. This continues up to a point, when a *ceiling effect* kicks in, and further changes in *X* have no subsequent effect on *Y*. This type of curve probably better typifies the relationship between job satisfaction and the number of days of work missed rather than the simple linear relationship shown earlier. At very high levels of satisfaction, days of attendance probably "max out" and people simply cannot attend work more, even if they became more satisfied with their jobs. At very low levels of satisfaction, there probably is a point where the dissatisfied worker has missed as many days as he or she can miss without being fired, so yet lower levels of satisfaction probably "bottom out" at this point. Whereas a linear model assumes that changes in satisfaction always lead to changes in the number of days missed, the S-shaped relationship recognizes that floor and ceiling effects operate.

In addition to an S-shaped relationship, variables can reflect an inverse-S-shaped relationship, which is illustrated in Figure 6.11. This might characterize, for example, the relationship between age in the elderly and the ability to recall a list of words. Elderly

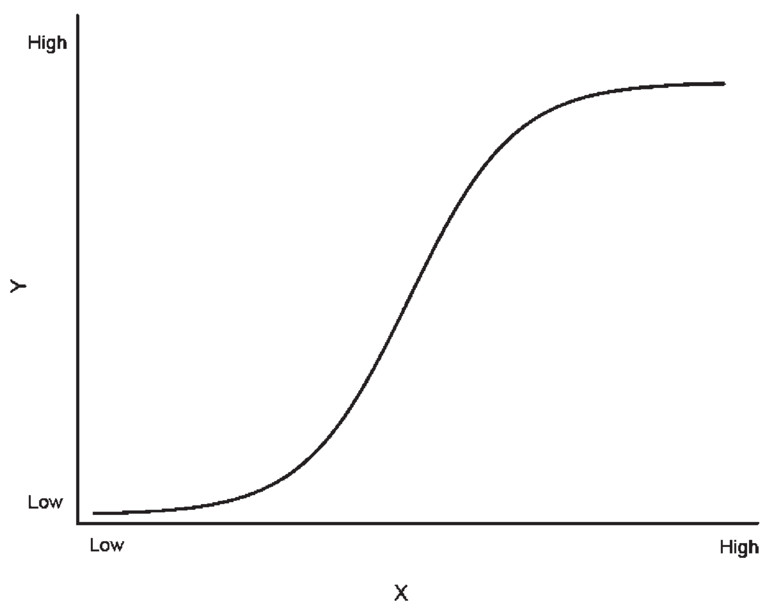


FIGURE 6.10. S-shaped function.

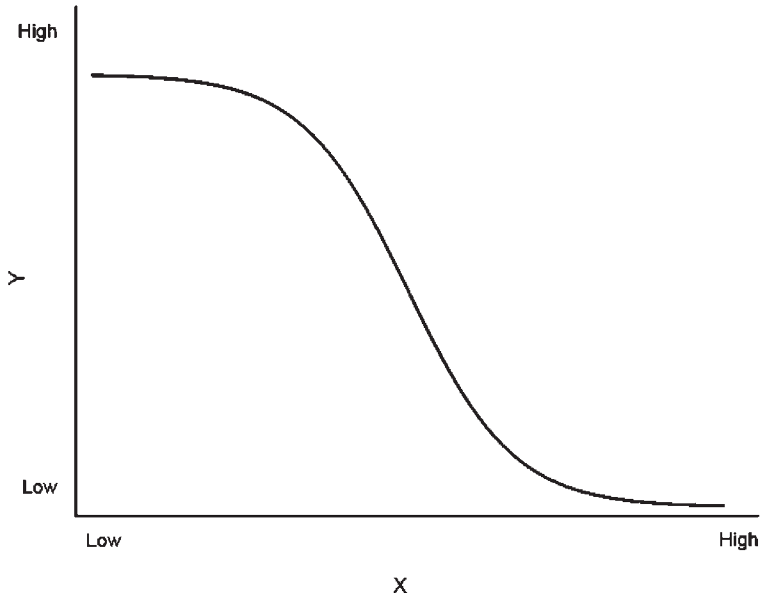


FIGURE 6.11. Inverse-S-shaped function.

people who are on the younger side are more likely to recall the list perfectly. As people get older, their ability to recall the list may start to decline until they reach an age where none of them can recall much of the list. Just as linear relationships can be direct or inverse, so can many forms of nonlinear relationships.

When Nonlinear Relationships Are Linear Relationships

Social scientists typically think in terms of linear relationships and posit and test their theories in ways that reflect this type of relationship. There are several reasons for this preference, some of which are defensible and others of which are not.

One dubious reason is that a family of statistical techniques that assumes linear relationships, known as linear regression and correlation, enjoys widespread use by social scientists. Thus, theorists tend to think in terms of linear relationships because their statistical tools focus on linear relationships. Granted, linear regression methods can be adapted to focus on nonlinear relationships (see Chapter 9), but, by far the most common use of regression methods is to test for, and impose, linear relationships on data. Theorists have become too accustomed to thinking in these terms, and perhaps some nonlinear thinking is necessary.

A second reason that theorists tend to ignore nonlinear relationships stems from a general preference for parsimony. The idea is that even though the true function between *X* and *Y* might be nonlinear, the departures from linearity are not sufficient enough to care about, and it is easier and more parsimonious to describe relationships in linear terms. Examine, for example, the S-shaped relationship in Figure 6.12, onto which a linear relationship has been superimposed. Note that for both relationships, low scores on *X* are associated with low scores on *Y*, moderate scores on *X* are associated with moderate scores on *Y*, and high scores on *X* are associated with high scores on *Y*. Granted there are floor and ceiling effects operating, but, the argument goes, these are of minor consequence in terms of the bigger trends characterizing the relationship. This argument has some validity as long as the departures from linearity truly are inconsequential. If, however, one's theory focuses on the very low or very high values of *X*, then it may be critical to take the nonlinearity into account. For example, suppose there is an S-shaped relationship between self-esteem (*X*) and the depressive symptom of lethargy (*Y*). An intervention might be developed to raise the self-esteem of people with low self-esteem on the assumption that such changes will reduce lethargy. But, it might be found instead that changes in self-esteem at the very low end of *X* have no effect on lethargy, given the S-shaped relationship. Because you assumed a linear relationship, you operated under the false impression that changes in self-esteem will have a positive effect on the expression of lethargy, even for those with very low self-esteem. The result would be a waste of resources and people receiving an ineffective treatment.

A third reason that theorists ignore nonlinear relationships is based on population partitioning. Suppose that the true relationship between *X* and *Y* is S-shaped but that the theorist is focused only on a population whose *X* scores occur between the dotted lines in Figure 6.13. For these individuals, the relationship between *X* and *Y* is indeed linear. So it makes sense to theorize accordingly, as long as the theorist does not generalize beyond the particular *X* values studied.

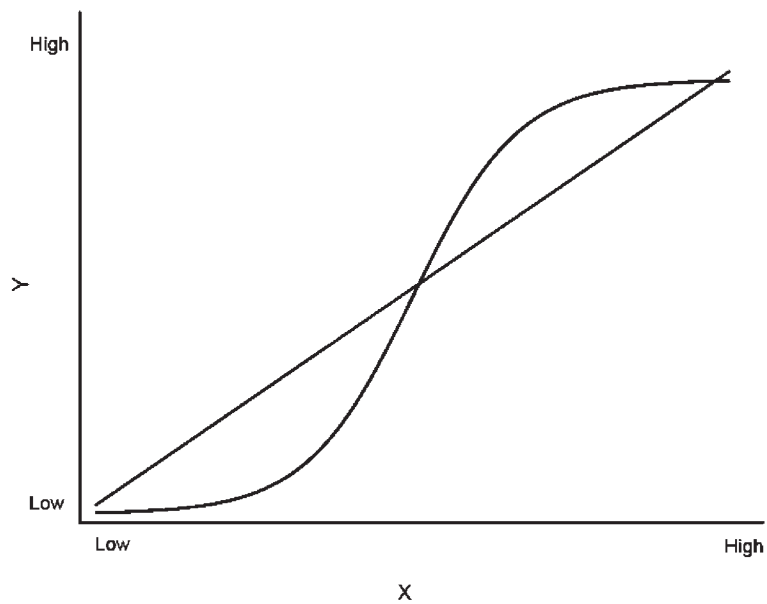


FIGURE 6.12. Linear and S-shaped functions.

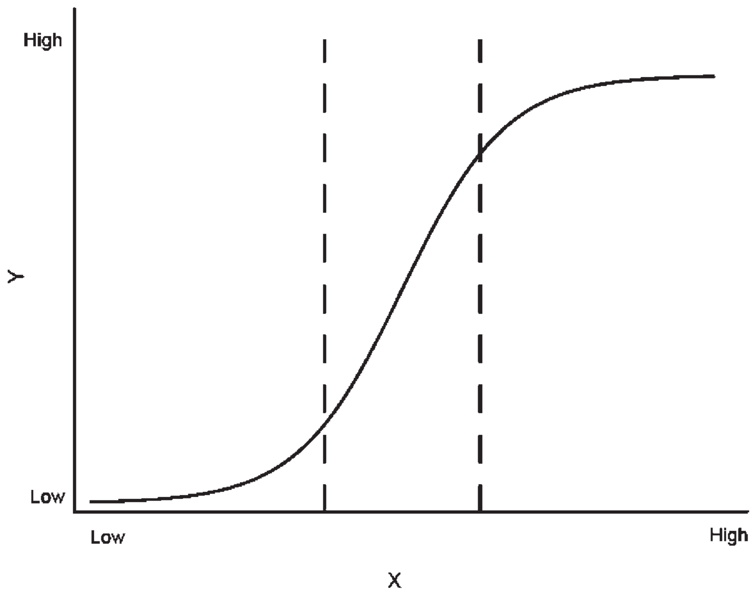


FIGURE 6.13. Focus on part of an S-shaped function.

A Thought Experiment with Hypothetical Scatterplots

Whenever you posit a relationship between two quantitative variables, you should "clarify the relationship" by drawing a shorthand hypothetical scatterplot for the two variables in the context of a thought experiment. Use linear functions if you think the relationship is linear and nonlinear functions if you think the relationship is nonlinear. Indicate the size of the effect by how steeply you draw your slope. Common nonlinear relationships to consider include the U-shaped relationship, the inverted-U-shaped relationship, the J-shaped relationship, the inverted-J-shaped relationship, the S-shaped relationship, the inverted-S-shaped relationship, a threshold function (where scores on *Y* are unchanged as *X* increases at low levels until a threshold point is reached and then a dramatic change in *Y* occurs at that threshold point), and various forms of logarithmic relationships (see Chapter 9). Figure 6.14 illustrates some of these relationships, emphasizing ones that we have not already graphed or that may not be obvious. We consider nonlinear relationships in more detail in Chapter 9.

As you think about and ultimately choose the "shape" of the curve describing the relationship between your quantitative variables, be sure that you can articulate the logic underlying why you think that particular function, and not others, operates. Sometimes you won't be sure about the curve that is most reasonable to posit. You may be able to narrow it to a few possible curves (e.g., direct linear or an S-shaped relationship), but you might not be able to choose between them on conceptual grounds. In this case, it is fine to use one of them as a "working assumption" and then develop your theory around it. But you also should explicitly recognize that other functions may be operating. The

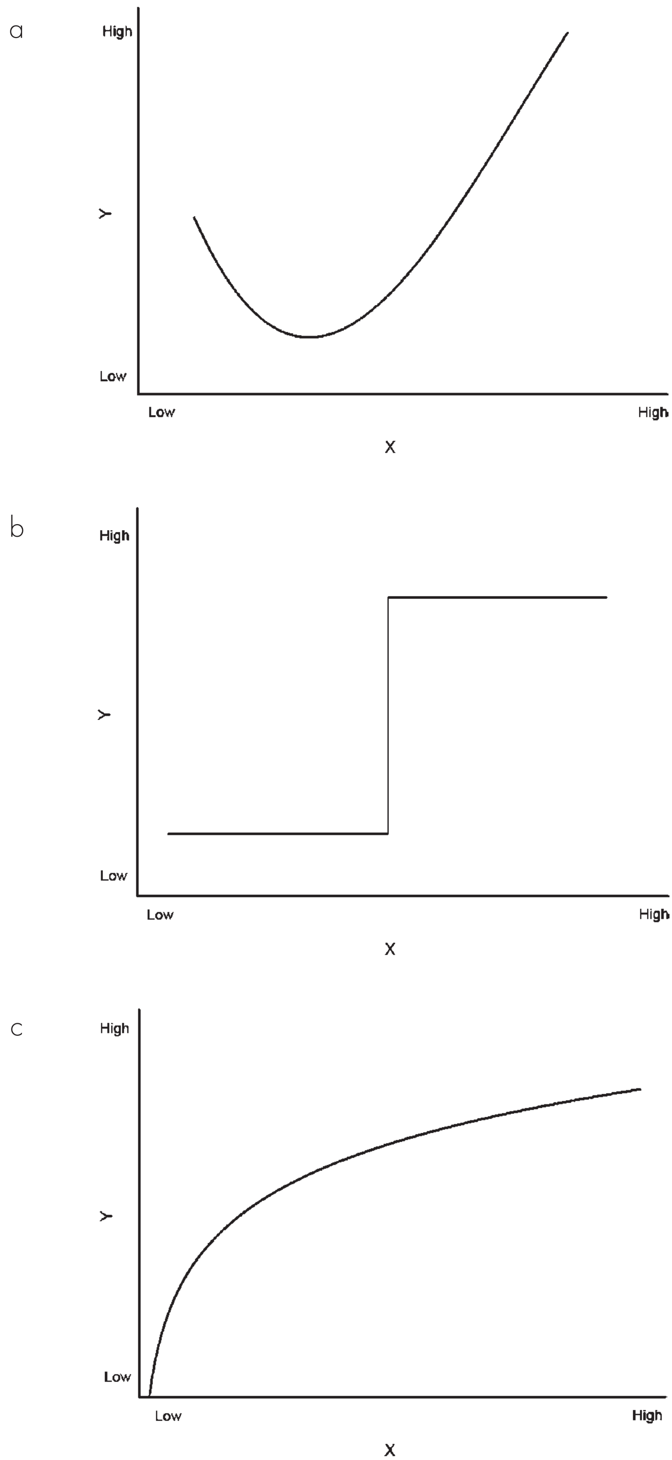


FIGURE 6.14. Additional examples of nonlinear functions. (a) J-shaped function; (b) threshold function; (c) logarithmic function.

default "working assumption" usually is a linear function because of its parsimony and familiarity.

When you describe your theory to others, you will need to describe the relationship you are positing. As with the hypothetical contingency table, typically you will not present the hypothetical scatterplot you used to focus the relationship. Again, it is a heuristic used in a thought experiment to help you think through and focus your presumed relationships. It is rare for social science theories to state the mathematical function they are assuming in the way we have done here. Rather, they simply state relationships in somewhat loose terms, such as "higher scores on intelligence are associated with higher grade-point averages." We think you should set your standards higher. To the extent possible, be explicit about the function, such as "higher scores on intelligence have a direct linear relationship with grade-point average."

In sum, when clarifying the relationship between two quantitative variables, you can do so by conducting a thought experiment that draws a shorthand hypothetical scatterplot. After drawing the scatterplot, state the relationship in formal, narrative terms and then articulate your underlying logic. Doing so will help you clearly explicate the form of the relationship. Again, try to articulate the logic behind the particular curve you draw.

THOUGHT EXPERIMENTS FOR RELATIONSHIPS BETWEEN CATEGORICAL AND QUANTITATIVE VARIABLES

When one of the variables is categorical and the other is quantitative, we suggest two different thought experiments to help focus the relationship: one using hypothetical means and the other using shorthand hypothetical scatterplots. The former is used when the "cause" is categorical and the "effect" is quantitative. The latter is used when the "cause" is quantitative and the "effect" is categorical. We consider each approach in turn.

Thought Experiments for a Categorical Cause and a Quantitative Effect: The Use of Hypothetical Means

When the categorical variable is the cause and the quantitative variable is the effect, the thought experiment we recommend uses hypothetical arithmetic means. We use a simple example to illustrate the method. Consider, as an example, the proposition that high school teachers at private schools have larger annual salaries than high school teachers at public schools. This leads us to generate the following table of hypothetical means

	Mean annual income
Private	45,000
Public	40,000

whereby the “cause” is listed as rows and the mean scores appear in the column. The relationship is clear: Teachers in private schools, on average, are thought to be paid more than teachers in public schools. This example is mundane, and you probably do not need to resort to a formal thought experiment to generate and/or clarify the theoretical proposition. Having said that, performing the exercise provides you with a sense of the size of the effect you expect, and as you think about this, it could generate ideas and clarify your logic. We consider more complex applications of hypothetical means later in this chapter.

As another example, consider a three-level categorical variable. For this example, the presumed cause is religious affiliation and the presumed effect is parental attitudes toward parental notification policies about adolescents seeking an abortion. This is a controversial policy. Some states have passed laws stating that if an adolescent under the age of 18 seeks an abortion, then the clinic or hospital that is to perform the abortion must notify a parent or legal guardian. Some people support this policy whereas others oppose it. The attitude is a quantitative variable that ranges from unfavorable to favorable, but it has no natural metric. Whereas “annual salary” in our previous example could naturally be considered in units of dollars, what are the natural units for an attitude? The answer is that there are none. To deal with this problem, we create arbitrary units ranging from 0 to 10, with 5 being a middle, neutral, or “moderate” attitude, numbers lower than 5 representing increasingly more negative attitudes, and numbers greater than 5 representing increasingly more positive attitudes. If your research uses a measure whose metric is well known, then you use that metric. However, in the absence of a natural or widely accepted metric, you can use an arbitrary scale from 0 to 10 or some other range.

Here is the table we create:

	Attitude
Catholics	7.0
Protestants	5.0
Jews	4.0

The theoretical proposition that derives from this table is as follows:

> **Proposition 1:** Religious affiliation is associated with the attitude toward parental notification such that Catholics have, on average, more positive attitudes than Protestants who, in turn, have more positive attitudes, on average, than Jews.

As with the hypothetical contingency table method, the theoretical proposition is specified by focusing on the column of hypothetical means and then stating the trends in the assigned numbers across the different row categories.

Thought Experiments for a Quantitative Cause and Categorical Effect: The Use of Hypothetical Probabilities

Consider the theoretical proposition that the degree to which a person is liberal affects his or her political party identification. Political party identification is a categorical variable (e.g., Democrat vs. Republican), whereas liberalness is a quantitative variable that varies from low to high on an underlying continuum. In such cases, two different types of thought experiments can be used to clarify the relationship. One thought experiment uses an extension of the hypothetical contingency table strategy for categorical variables; it is presented in Appendix 6A at the end of this chapter. The second thought experiment uses a combination of the hypothetical contingency table and the hypothetical scatterplot to clarify the relationship. We consider this strategy here, although you might find the thought experiment in the appendix to be more to your liking. We begin by examining the case of a two-level categorical variable and then we consider categorical variables with three or more levels.

For this thought experiment, we must turn the categorical variable into a quantitative representation so that we can draw a scatterplot between two "quantitative" variables. The *X* axis of the scatterplot describes values of the "cause" or, in this case, liberalness. On the *Y* axis, we will plot the probability that people with a given liberalness score are Democrats. To make this relationship concrete, we begin by arbitrarily creating a metric for the liberalness dimension ranging from 1 to 5, where individuals with a score of 5 are thought to be extremely liberal, individuals with a score of 4 are quite liberal, individuals with a score of 3 are moderately liberal, individuals with a score of 2 are slightly liberal, and individuals with a score of 1 are not at all liberal. We then create a hypothetical contingency table using the "cause" as rows and the "effect" as columns, much like our previous case with two categorical variables:

	Democrat	Republican	Total
Liberalness = 5			100
Liberalness = 4			100
Liberalness = 3			100
Liberalness = 2			100
Liberalness = 1			100

Next, we fill in the numbers that we think capture the relationship, just as we did in the contingency table method. For example, we might fill in the following numbers:

	Democrat	Republican	Total
Liberalness = 5	70	30	100
Liberalness = 4	60	40	100

Liberalness = 3	50	50	100
Liberalness = 2	40	60	100
Liberalness = 1	30	70	100

A cell entry reflects the percentage of people with a given liberalness score whom we think will be Democrats or Republicans. We can convert these percentages to probabilities by dividing entries by 100:

	Democrat	Republican
Liberalness = 5	.70	.30
Liberalness = 4	.60	.40
Liberalness = 3	.50	.50
Liberalness = 2	.40	.60
Liberalness = 1	.30	.70

We now arbitrarily select one column on which to focus, and it will be the Democrats (the choice is arbitrary because the probability for Democrats is the mirror image of the probability for Republicans). We now construct a scatterplot of the relationship (see Figure 6.15) between the liberalness scores and the probability of being a Democrat (using the entries in the Democrat column). Note that the relationship is direct linear in form.

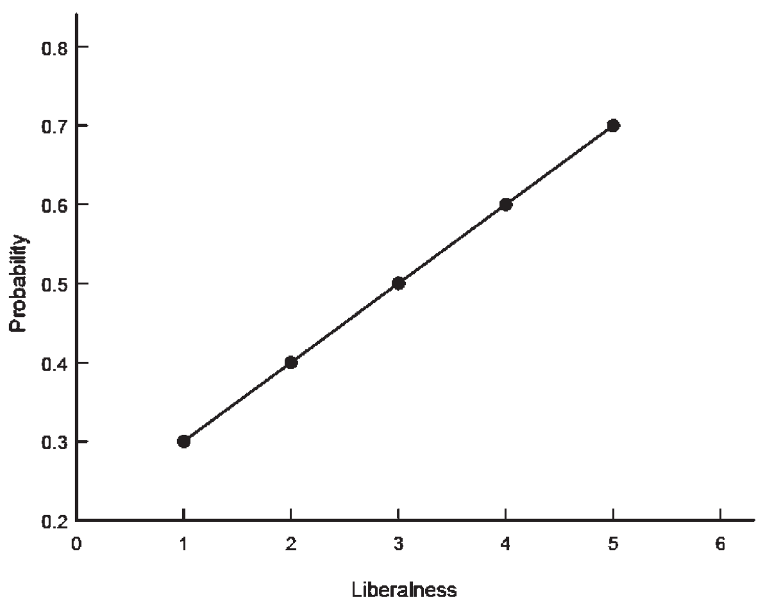


FIGURE 6.15. Liberalness example.

But perhaps instead of direct linear, the relationship between liberalness and the probability of being a Democrat is S-shaped. Indeed, any of the nonlinear relationships we discussed for two quantitative variables could be used. What we are suggesting is applying the shorthand hypothetical scatterplot method to two variables: (1) the quantitative cause (*X*) and (2) the probability of being in a given category of the categorical variable (in this case, being a Democrat). For the scatterplot in Figure 6.15, the theoretical proposition that results is

> **Proposition 1:** The probability of being a Democrat is a direct linear function of liberalness such that as liberalness increases, the probability of being a Democrat increases.

To use this heuristic, you need not generate the probabilities that we used in constructing Figure 6.15. We did this for pedagogical reasons. All you need to do is create a shorthand scatterplot that draws the function that you think relates the *X* variable to the "probability of being in category *y*" for the *Y* or outcome variable. Articulate your logic as you do this. When there are more than two categories for *Y*, you will need to describe the *X* and "probability of *y*" function for each *Y* category separately, because it is only in the two-category case that the one set of probabilities (e.g., for Democrats) is the mirror image of the other set of probabilities (e.g., for Republicans).

In sum, another tool that you can use for clarifying the relationship between a quantitative cause and a categorical effect is a shorthand scatterplot wherein the categorical variable is recast in the form of probability values (but see also the thought experiment in Appendix 6A as an alternative).

THOUGHT EXPERIMENTS FOR MODERATED RELATIONSHIPS

Thus far we have considered methods for clarifying bivariate relationships. There is another type of relationship that is important in theory construction, called a *moderated relationship*. Moderated relationships involve three variables and focus on cases where the strength or nature of a relationship between two variables changes depending on the value of a third variable. For example, we might theorize that inflation has a bigger influence on economies in underdeveloped countries as opposed to developed countries. Or we might theorize that higher levels of education are more likely to translate into job opportunities for whites as opposed to blacks. This section articulates the nature of moderated relationships and thought experiments for clarifying them.

Thought Experiments Using Hypothetical Factorial Design

To clarify a moderated causal relationship, we use a heuristic device called a *hypothetical factorial design*. Begin the task by identifying the cause, the effect, and the moderator variable. To make this concrete, consider the case wherein we examine how satisfied adolescents are with their relationships with their mothers as a function of gender. The outcome variable is the adolescent's relationship satisfaction with his or her mother (Y), and the presumed cause is gender (X). We introduce new terminology by referring to the presumed cause (gender) as the *focal independent variable*, a term that is typically used in the literature on moderated relationships.

We offer the theoretical proposition that girls will be more satisfied with their relationships with their mothers than boys, because girls have more in common with their mothers and can more easily identify with them. However, we also expect this gender difference to vary as a function of the age of the adolescent (Z). More specifically, we conjecture that adolescents start to grow apart from their mothers as they become older, and that this process is truer for boys than girls. We reason that as boys and girls go through the pubertal and physical changes with age, girls become more like their mothers physically and their interests probably begin to converge with those of their mothers as well (e.g., in clothes, shopping, and the like). By contrast, as boys become physically more mature, they separate more from their mothers, weakening the relationship between them. These dynamics should produce a greater gender difference in relationship satisfaction as adolescents grow older.

In this example, age of the adolescent is represented by his or her grade in school. To simplify matters, we use only two grade levels, seventh and eighth grades. Grade is a moderator variable: It is thought to moderate the strength of the relationship between gender and relationship satisfaction with the mother. At earlier grades, when adolescents are younger, we expect the gender difference in relationship satisfaction to be weaker than at later grades, when adolescents are older.

To focus and clarify the relationship, we construct a factorial table in which we list the levels of the moderator variable as columns and the levels of the focal independent variable as rows. Here is the table:

	Grade 7	Grade 8
Female		
Male		

Next we fill in hypothetical mean scores for the outcome variable to reflect the theoretical dynamics we believe operate (using the same logic as the hypothetical mean approach, discussed earlier). Because satisfaction has an arbitrary metric, we use a "scale" ranging from 0 to 10, with higher scores indicating higher levels of satisfaction. Here is the table with the numerical entries we made up:

	Grade 7	Grade 8
Female	6.0	6.0
Male	5.0	4.0

Let's examine this table more closely. The gender difference for seventh graders is represented by the mean difference between females and males in the first column. This difference is

$$M_{F7} - M_{M7} = 6.0 - 5.0 = 1.0$$

where M_{F7} is the hypothetical mean for females in grade 7 and M_{M7} is the hypothetical mean for males in grade 7. We can calculate a similar gender difference for eighth graders:

$$M_{F8} - M_{M8} = 6.0 - 4.0 = 2.0$$

If the gender differences are the same at both grade levels, then the value for the first difference should be identical to the value of the second difference. But this is not the case, as seen by differencing the two differences:

$$\begin{aligned} \text{IC} &= (M_{F7} - M_{M7}) - (M_{F7} - M_{M7}) \\ &= (6.0 - 5.0) - (6.0 - 4.0) \\ &= 1.0 - 2.0 = -1.0 \end{aligned}$$

where IC stands for an interaction contrast. We use the term *interaction contrast* because a moderated relationship maps onto an interaction effect in factorial designs. An interaction contrast in a factorial design is the comparison of the effect of a focal independent variable at one level of the moderator variable with the effect of the focal independent variable at another level of the moderator variable. The above table yields a nonzero interaction contrast, so a moderated relationship is implied by the numerical entries.

To recap, to clarify a moderated relationship:

1. Create a factorial table with the moderator variable as columns and the focal independent variable as rows.
2. Fill in plausible hypothetical mean values on the outcome variable for each cell of the table.
3. Calculate the effect of the focal independent variable at each level of the moderator variable and then calculate the interaction contrast to determine if there is a moderated relationship.

The following theoretical proposition results from the above table:

Proposition 1: Gender differences in adolescent relationship satisfaction during grade 8 are larger than gender differences in grade 7.

This proposition describes in words the "results" of the interaction contrast of hypothetical means. Of course, you need to be able to articulate the logic and reasons why you expect this proposition to be true, as we did at the outset of this example. You would think about this carefully as you assign your hypothetical mean scores in each cell of the factorial table.

A simple way of capturing the numerical computations for an interaction contrast in a factorial table with two rows and two columns (called a 2 × 2 table) is to label the cells as follows:

	Grade 7	Grade 8
Female	*a*	*c*
Male	*b*	*d*

Then calculate $(a - b) - (c - d)$ to obtain the value of the interaction contrast. If this value is nonzero, then a moderated relationship is present. If the value is zero, then there is no moderated relationship.

Hypothetical Factorial Designs with More Than Two Levels

Sometimes your variables may have more than two levels. For example, suppose that instead of grades 7 and 8, we had grades 7, 8, and 9 in our population, yielding the following table:

	Grade 7	Grade 8	Grade 9
Female			
Male			

Here is the table with hypothetical means filled in by the theorist:

	Grade 7	Grade 8	Grade 9
Female	6.0	6.0	6.0
Male	5.0	4.0	4.0

To elucidate the moderated relationship, we need to compare the gender difference at all possible pairs of levels of the moderator variable. The easiest way to do this is to construct all possible 2 × 2 subtables from the larger design. In this case, there are three of them. Here they are with the value of the 2 × 2 interaction contrast computed beneath each:

	Grade 7	Grade 8
Female	6.0	6.0
Male	5.0	4.0

(6 – 5) – (6 – 4) = –1

	Grade 7	Grade 9
Female	6.0	6.0
Male	5.0	4.0

(6 – 5) – (6 – 4) = –1

	Grade 8	Grade 9
Female	6.0	6.0
Male	4.0	4.0

(6 – 4) – (6 – 4) = 0

If any of the 2 × 2 tables has a nonzero interaction contrast, then a moderated relationship exists. In this example, grade moderates the effect of gender on satisfaction, but the description of the theoretical relationship is more complex than in the two-level case. Here are the theoretical propositions we derive from the tables:

> **Proposition 1:** Gender differences in adolescent relationship satisfaction during grade 8 are larger than gender differences in grade 7.
>
> **Proposition 2:** Gender differences in adolescent relationship satisfaction during grade 9 are larger than gender differences in grade 7.
>
> **Proposition 3:** Gender differences in adolescent relationship satisfaction during grades 8 and 9 are the same.

These propositions describe in words the "results" of the interaction contrasts in each of the three 2 × 2 subtables. Again, you need to be able to articulate the conceptual mechanisms underlying the numbers you assign in the table.

As another example with more than two levels, consider the case where the focal independent variable has more than two levels, as in the following table:

	Male	Female
Divorced		
Separated		
Married		

In this example, the outcome variable is again relationship satisfaction (Y), the moderator variable is gender (Z), and the focal independent variable is the marital status of the adolescent's mother (X). The theorist conjectures that the effect of marital status on relationship satisfaction varies as a function of gender. For boys, the satisfaction with the maternal relationship is thought to be unaffected by the marital status of the mother. For girls, relationship satisfaction is thought to be lower when there is no father figure in the home (separated and divorced) as opposed to when there is (married). The idea is that the presence of a father provides more resources for the family so that the mother can spend more free time with the daughter than would otherwise be the case. This, in turn, creates a closer bond between mother and daughter. Here is the table with hypothetical means inserted:

	Male	Female
Divorced	4.0	4.0
Separated	4.0	5.0
Married	4.0	6.0

We again form all possible 2 × 2 tables. They are:

	Male	Female
Divorced	4.0	4.0
Separated	4.0	5.0

(4 – 4) – (4 – 5) = 1

	Male	Female
Divorced	4.0	4.0
Married	4.0	6.0

(4 – 4) – (4 – 6) = 2

	Male	Female
Separated	4.0	5.0
Married	4.0	6.0

(4 – 4) – (5 – 6) = 1

If all of the interaction contrasts in the 2 × 2 subtables equal zero, then there is no moderated relationship. In this case, some of the contrasts are nonzero, so a moderated relationship exists. Specifically, the mean difference between adolescents with divorced versus separated mothers is larger for females than it is for males. Also, the mean difference between adolescents with divorced versus married mothers is larger for females than it is for males. Finally, the mean difference between adolescents with separated versus married mothers is larger for females than it is for males. Here are the theoretical propositions that emerge from the tables:

> **Proposition 1:** The effect of being divorced versus separated is stronger for females than for males.
>
> **Proposition 2:** The effect of being divorced versus married is stronger for females than for males.
>
> **Proposition 3:** The effect of being separated versus married is stronger for females than for males.

Note that each proposition focuses on a different 2 × 2 subtable and its corresponding interaction contrast. As always, you need to be able to articulate the mechanisms that justify the numerical assignments you make.

Hypothetical Factorial Designs with Quantitative Variables

To focus a moderated relationship in which a quantitative variable is involved as either the focal independent variable or the moderator variable, you still form a factorial table of hypothetical means, but now you group the values of the quantitative variable into "high," "medium," and "low" categories.[1] As an example, consider the case where the moderator variable is gender of the adolescent (*Z*), the focal independent variable is how much time a mother spends with her child (*X*), and the outcome variable is relationship satisfaction with the mother (Y). The theorist conjectures that the time spent with boys will not impact relationship satisfaction, but it will do so for girls. The presumed mechanism is that for girls, the more time the mother and daughter spend together, the more the daughter will identify with the mother, leading to a closer relationship. By contrast, boys will not identify with the mother as much, even with increased time together. The factorial table with the hypothetical means might appear as follows:

	Male	Female
High	3.0	4.0
Medium	3.0	5.0
Low	3.0	6.0

You calculate all possible 2 × 2 subtables, as before. They are:

	Male	Female
High	3.0	4.0
Medium	3.0	5.0

(3 – 3) – (4 – 5) = 1

	Male	Female
High	3.0	4.0
Low	3.0	6.0

(3 – 3) – (4 – 6) = 2

	Male	Female
Medium	3.0	5.0
Low	3.0	6.0

(3 – 3) – (5 – 6) = 1

These tables lead to the following theoretical propositions:

Proposition 1: The effect of spending relatively large amounts of time together versus more moderate amounts of time together on relationship satisfaction is stronger for females than for males.

Proposition 2: The effect of spending relatively large amounts of time together versus small amounts of time together on relationship satisfaction is stronger for females than for males.

Proposition 3: The effect of spending relatively moderate amounts of time together versus small amounts of time together on relationship satisfaction is stronger for females than for males.

Each proposition describes in words the "results" of the interaction contrasts for each 2 × 2 subtable. As before, you would elaborate the mechanisms underlying each proposition.

Hypothetical Scatterplots and Quantitative Variables

There is a different way of conceptualizing the above moderated relationship that takes advantage of the fact that both X and Y are quantitative variables. Because each is quan-

[1]Although we group quantitative variables for theory construction heuristics in hypothetical factorial designs, this by no means suggests we would do so when analyzing data. This is a bad practice for purposes of data analysis.

titative in nature, we can draw a hypothetical scatterplot between them at selected levels of the moderator variable—in this case, one slope for males and one for females. This has been done in Figure 6.16. Note that although both relationships are linear, the line is steeper for females than it is for males, suggesting the relationship between *X* and *Y* differs for the two groups. A way of characterizing this in terms of a theoretical proposition is as follows:

> **Proposition 1:** The direct linear relationship between time spent together and relationship satisfaction is stronger for females than males.

In this case, the focus is on describing the differences in the slopes of the two scatterplots. As before, you must be able to articulate the mechanisms that underlie these slope differences.

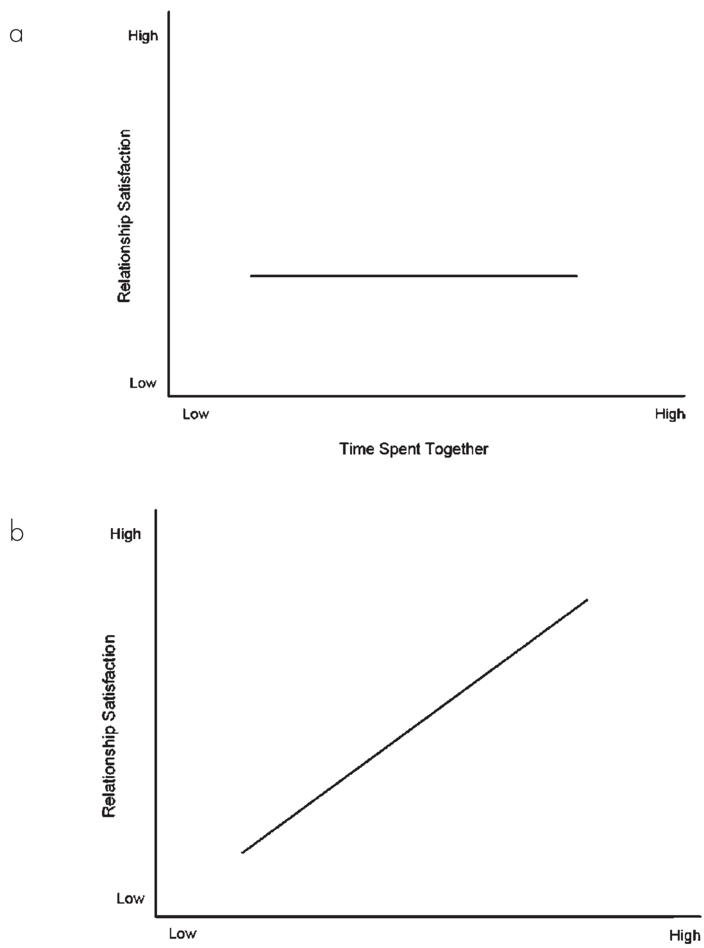


FIGURE 6.16. Hypothetical scatterplots for moderated relationship. (a) Males; (b) females.

Summary for Moderated Relationships

Moderated relationships are a part of some theories, and it is important to be clear about the nature of a moderated relationship. Moderated relationships involve scenarios where the strength or nature of a relationship between two variables is thought to vary as a function of a third (moderator) variable. Moderated relationships can be incorporated into a theory by asking questions about whether the effect of *X* on *Y* will be equally strong in all circumstances or whether the effect of *X* on *Y* will be equally strong for all individuals—that is, using what we call the *stronger-than heuristic*. This heuristic involves asking if the effect of *X* on *Y* will be stronger in some circumstances than in others or stronger for some individuals than for others. If the answer is yes, then a moderated relationship is suggested. A useful strategy for clarifying a moderated relationship is a thought experiment with a hypothetical factorial design. More complex moderated relationships can be offered in a theory, an example of which is described in Appendix 6B. We discuss moderated relationships in more detail in Chapter 7.

BROADER USES OF HYPOTHETICAL FACTORIAL DESIGNS IN THOUGHT EXPERIMENTS

Sometimes you will embark on the theory construction process with only a rough, intuitive sense about the relationship of two different "causes" to an effect. You may not be able to articulate your logic well. In such cases, it might help to construct a factorial table of hypothetical means and fill in the numbers based on your intuition. After doing so, the numbers in the table can be used to make the relationships you have intuitively generated more explicit. For example, suppose we examine the effects of religion (Catholic vs. Protestant) and religiosity on attitudes toward the complete legalization of abortion. Because religiosity is a continuous variable, we group it into the three categories of low, medium, and high. Because there is no natural metric for attitudes toward legalizing abortion, we use a 0–10 scale for it, with higher scores indicating a more favorable attitude. We create the following table

	Religiosity		
	Low	Medium	High
Catholic			
Protestant			

and then fill in hypothetical means:

	Religiosity		
	Low	Medium	High
Catholic	7.0	5.0	3.0
Protestant	7.0	5.0	6.0

Three classes of relationships can be derived from this table: one based on main effects, one based on simple main effects, and one based on interaction contrasts. To illustrate them, we revise the table by adding marginal means to it, as follows:

	Religiosity: Low	Medium	High	
Catholic	7.0	5.0	3.0	5.0
Protestant	7.0	5.0	6.0	6.0
	7.0	5.0	4.5	

A marginal mean is the average of numbers in the respective row or column associated with that marginal mean. The marginal mean for low religiosity is (7.0 + 7.0)/2 = 7.0. The marginal mean for medium religiosity is (5.0 + 5.0)/2 = 5.0. The marginal mean for high religiosity is (3.0 + 6.0)/2 = 4.5. The marginal mean for Catholics is (7.0 + 5.0 + 3.0)/3 = 5.0. The marginal mean for Protestants is (7.0 + 5.0 + 6.0)/3 = 6.0.

Relationships Characterized by Main Effects

The main effects in the factorial table of hypothetical means focus on the marginal means for a given variable. For example, the main effect of religiosity focuses on the marginal means for religiosity. Looking at these means, it can be seen that we posit that higher scores on religiosity are associated with less positive attitudes toward abortion (low religiosity = 7.0, medium religiosity = 5.0, and high religiosity = 4.5). Thus, the relationship is inverse. However, note that we have posited a nonlinear relationship with the numbers we assigned. This is because the difference between low religiosity and medium religiosity (7.0 – 5.0 = 2.0) is larger than the difference between medium religiosity and high religiosity (5.0 – 4.5 = 0.5). Shifts in religiosity at the lower end of the construct (from low to medium) have larger effects on attitudes toward abortion than comparable shifts (from medium to high) at the upper end of the construct. As we think through the logic of this "finding," we might change our minds about the cell entries we have made and revise them so that this relationship is linear. For this example, we will leave them as is and posit a nonlinear relationship, leading to the following proposition:

> **Proposition 1:** Religiosity is inversely related to attitudes toward abortion such that higher scores on religiosity tend to be associated with more negative attitudes toward abortion. However, the relationship is nonlinear, with greater shifts in attitudes as a function of religiosity at the low levels of religiosity as compared to high levels of religiosity.

As we reflect on this proposition and the numbers we have entered, we try to articulate the logic that might justify it. It is not enough to just state a proposition. Rather, there must be some underlying logic that convinces the reader that it is viable. In this case, we might argue that both Catholicism and Protestantism formally discourage abortion, so that individuals who identify with or embrace those religions (as reflected by religiosity) should be more likely to oppose the complete legalization of abortion. The effect of increased identification with these religions, however, has its limits, and there comes a point where increased religiosity has diminishing incremental effects on opposition to abortion—hence the nonlinear relationship.

We repeat this process examining the main effect for the second variable, in this case, religion. The marginal mean for Catholics is 5.0 and the marginal mean for Protestants is 6.0. This leads to our second proposition:

> **Proposition 2:** Catholics, on average, have more negative attitudes toward abortion than Protestants.

Again, we want to try to articulate the logic for this proposition, perhaps by trying to build a case that the formal teachings of Catholicism about abortion are more extreme or more salient to individuals than those of Protestantism.

Relationships Characterized by Simple Main Effects

The second class of relationships focuses on what is called *simple main effects*. The best way to conceptualize these types of effects is to assign one of the variables the role of the focal independent variable and the other the role of the moderator variable. This choice is arbitrary and depends on how we want to frame our theory, a point we return to shortly. For this example, we declare religion to be our focal independent variable and religiosity to be our moderator variable.

A *simple main effect* refers to the effect of the focal independent variable at a given level of the moderator variable. For example, looking at the factorial table, we see that at low levels of religiosity, there is no difference, on average, in attitudes toward abortions for Catholics and Protestants: $(M_{CL} - M_{PL}) = 7 - 7 = 0$. This also is true at medium levels of religiosity: $(M_{CM} - M_{PM}) = 5 - 5 = 0$. However, at high levels of religiosity, there is a difference between Catholics and Protestants, $(M_{CH} - M_{PH}) = 3 - 6 = -3$, with Catholics being more negative toward abortion than Protestants. These characterizations, called *simple main effects*, reflect the effect of a focal independent variable at a given level of the moderator variable. In this case, the three simple main effects lead to the following propositions:

> **Proposition 3:** When religiosity is low, attitudes toward abortion for Catholics will be the same as attitudes toward abortion for Protestants, on average.
>
> **Proposition 4:** When religiosity is moderate, attitudes toward abortion for Catholics will be the same as attitudes toward abortion for Protestants, on average.

Proposition 5: When religiosity is high, attitudes toward abortion for Catholics will be less favorable than attitudes toward abortion for Protestants, on average.

The theorist would then articulate the logic underlying each of these propositions.

Relationships Characterized by Interaction Contrasts

Finally, we can specify the interaction contrasts, which compare the effects of the focal independent variable at one level of the moderator variable with the effects of that focal independent variable at another level of the moderator variable. This process is best captured by calculating all possible 2 × 2 subtables, as before. They are:

	Low	Medium
Catholic	7.0	5.0
Protestant	7.0	5.0

(7 – 7) – (5 – 5) = 0

	Low	High
Catholic	7.0	3.0
Protestant	7.0	6.0

(7 – 7) – (3 – 6) = –3

	Medium	High
Catholic	5.0	3.0
Protestant	5.0	6.0

(5 – 5) – (3 – 6) = –3

These tables lead to the following theoretical propositions (focusing on each subtable):

Proposition 6: Catholic–Protestant differences in attitudes toward abortion are the same at low levels of religiosity as they are at medium levels of religiosity.

Proposition 7: Catholic–Protestant differences in attitudes toward abortion are smaller at low levels of religiosity than they are at high levels of religiosity.

Proposition 8: Catholic–Protestant differences in attitudes toward abortion are smaller at medium levels of religiosity than they are at high levels of religiosity.

This thought experiment, using intuitive entries in the factorial table of hypothetical means, generates many interesting theoretical propositions. As you work through each proposition, you should think about the logic that underlies it and whether the relationship in question is logically defensible. Articulate that logic. As you pursue this process, you might revise your entries to better conform to the logic that emerges from your thought experiment. The process is one of “back-and-forth” revisions between the table entries and theory, until you settle upon a pattern of entries and a set of theoretical propositions that are coherent and logical.

Sometimes you might be able to generate reasonable logic for several patterns of entries. This is of particular interest because then you have specified competing theories. One set of logical thoughts suggests that the entries should pattern themselves one way and another set of logical thoughts suggests that the entries should pattern themselves a different way. Which pattern of results occurs in practice? The best way to find this out is to conduct a formal study (not a thought experiment) and explore the matter empirically.

Choice of the Moderator Variable

When two variables, *X* and *Z*, form a moderated relationship to impact *Y*, the designation of which variable takes the role of moderator is often straightforward. For example, suppose you want to determine if the effectiveness of a clinical treatment for depression is more effective or less effective for males as compared with females. It is clear in this case that gender is the moderator variable and the presence or absence of the treatment is the focal independent variable. On the other hand, there are situations in which one theorist's moderator variable might be another theorist's focal independent variable, and vice versa. For example, a consumer psychologist who studies product quality and choice might be interested in the effect of product quality on brand preference and how this is moderated by pricing structure. In contrast, a marketing researcher might be interested in the effect of pricing structure on brand preference and how this is moderated by product quality. In both cases, the designation of the moderator variable follows directly from the theoretical orientation of the researcher. Neither specification is better than the other. The two designations simply represent different perspectives on the same phenomenon. You choose a variable to take the role of the moderator based on how you want to frame your theory. In the example in the previous sections, we chose to frame the theory in terms of how religiosity moderates the effect of religious denomination on attitudes toward abortion.

SUMMARY AND CONCLUDING COMMENTS

When constructing a theory, in addition to clearly articulating the constructs in that theory, you must clearly articulate the (expected) relationships between constructs. Whereas Chapter 5 presented heuristics for focusing concepts, this chapter presented strategies for clarifying theoretical relationships in the form of thought experiments. These included the use of hypothetical contingency tables, hypothetical mean tables, shorthand hypothetical scatterplots, hypothetical probability scatterplots, and hypothetical factorial designs. The approaches we discussed are not exhaustive, and we illustrate others in later chapters. However, if you carefully scrutinize every relationship in your theory using thought experiments, and try to articulate the logic and reasoning underlying the proposed relationships, you will be that much further along in the process to conceiving a well-thought-out theoretical framework.

SUGGESTED READINGS

Jaccard, J., & Becker, M. (2001). *Statistics for the behavioral sciences.* Belmont, CA: Wadsworth.—An introductory statistics text describing scatterplots, contingency tables, means, and factorial designs.

Pearl, J. (2000). *Causality: Models, reasoning, and inference.* New York: Cambridge University Press.—A discussion of relationships focusing on causality. Includes a discussion of Simpson's paradox, which is in Box 6.1.

KEY TERMS

thought experiment (p. 91)
categorical variable (p. 93)
quantitative variable (p. 94)
contingency table (p. 94)
scatterplot (p. 98)
direct linear relationship (p. 101)
inverse linear relationship (p. 101)
zero-slope relationship (p. 101)
shorthand hypothetical scatterplot (p. 102)
inverted-U relationship (p. 102)
S-shaped relationship (p. 105)
floor effect (p. 105)
ceiling effect (p. 105)
population partitioning (p. 107)
hypothetical mean heuristic (p. 110)
hypothetical probabilities (p. 112)
moderated relationships (p. 114)
focal independent variable (p. 115)
moderator variable (p. 116)
interaction contrast (p. 116)
stronger-than heuristic (p. 122)
main effect (p. 123)
simple main effect (p. 124)

EXERCISES

Exercises to Reinforce Concepts

1. What is the difference between a categorical and a quantitative variable?
2. Describe the hypothetical contingency table approach to focusing the relationship between two categorical variables. Include in your discussion how you would derive theoretical propositions from the table.
3. Describe a scatterplot.
4. Describe the hypothetical scatterplot heuristic for focusing a relationship between quantitative variables.
5. Draw a scatterplot that shows a positive direct linear relationship; draw one that shows an inverse linear relationship.
6. Draw a graph with two linear relationships on it, but one with a stronger effect as reflected by the slopes.
7. Describe how two variables that are nonlinearly related can exhibit a linear relationship.

BOX 6.1. Simpson's Paradox

Scientists have discovered an interesting phenomenon with respect to relationships between variables, called *Simpson's paradox*. Simpson's paradox refers to the reversal of the direction of an association when data from several groups are combined to form a single group. It is a surprising finding that shows that we must be careful when aggregating results across populations or studies to characterize relationships. Consider the following three scenarios, each of which illustrates the paradox:

Scenario 1

A study examined death rates from tuberculosis in two cities, Richmond, Virginia, and New York City, in 1910. They found the following:

1. The death rate for African Americans was lower in Richmond than in New York City.
2. The death rate for European Americans was lower in Richmond than in New York City.
3. Even though the death rate was lower for African Americans and for European Americans in Richmond than in New York City, the death rate for the combined population of African Americans and European Americans was *higher* in Richmond than in New York City.

Scenario 2

Psychological researchers test two treatments for a mental illness in two separate clinical trials.

1. In trial number 1, treatment A cures 20% of its cases (40 out of 200) and treatment B cures 15% of its cases (30 out of 200). So, treatment B is inferior to treatment A, because its cure rate is lower than treatment A's.
2. In trial number 2, treatment A cures 85% of its cases (85 out of 100) and treatment B cures 75% of its cases (300 out of 400). So, treatment B is inferior to treatment A, because its cure rate is lower than treatment A's.
3. Pooling data across the two trials, treatment A cured 125 out of 300 cases (or 45%) and treatment B cured 330 out of 600 cases (or 55%). So, treatment B is *superior* to treatment A, because its cure rate is higher than treatment B's.

Scenario 3

Suppose that a university is trying to discriminate in favor of women when hiring faculty to atone for past hiring bias. It advertises positions in Department A and in Department B and only in those departments.

1. Five men apply for the positions in Department A and one is hired. Eight women apply and two are hired. The hiring rate for men is 20% and

cont.

for women it is 25%. Department A thus has favored hiring women over men.

2. For Department B, eight men apply and six are hired, and five women apply and four are hired. The hiring rate for men is 75% and for women it is 80%. Department B also has favored hiring women over men.
3. If we pool the data across the university as a whole, 13 men and 13 women applied for jobs, and 7 men and 6 women were hired. The hiring rate for men is higher than the hiring rate for women.

Why does Simpson's paradox occur? Does it have something to do with the unequal sample sizes in the scenarios? The answer is, not really. For example, in Scenario 3, there were 13 male and 13 female applicants, namely, equal sample sizes for both genders. Department A had 13 applicants, as did Department B, again yielding equal sample sizes. How about the fact that the sample sizes were small? This also is not the source of the paradox. If you multiply all the numbers by 1,000 in Scenario 3, for example, the paradox still exists.

Simpson's paradox has a mathematical explanation. For eight whole numbers, represented by the letters *a, b, c, d, A, B, C, D,* it is possible to have the following relationships

$$a/b < A/B$$
$$c/d < C/D$$
$$\text{and } (a + c)/(b + d) > (A + C)/(B + D)$$

which is what occurs in Simpson's paradox. For example:

$$1/5 < 2/8$$
$$6/8 < 4/5$$
$$\text{yet, } 7/13 > 6/13$$

Simpson's paradox emerges in the three scenarios because a subtle bias is operating, which we explicate using the hiring example in Scenario 3. In this example, more women are applying for jobs that are harder to get and more men are applying for jobs that are easier to get. For example, it is harder to be hired in Department A than in Department B, and more women are applying to Department A than to Department B. Simpson's paradox occurs between either of the two following extremes: (1) when slightly more women are applying for jobs that are much harder to get, and (2) when many more women are applying for jobs that are slightly harder to get.

When social scientists aggregate across groups of people, they must be careful about the causal inferences they make, as Simpson's paradox illustrates. Simpson's paradox has implications for commonly used methods of research synthesis that aggregate findings across studies, such as the quantitative method of meta-analysis.

8. What are some of the reasons why a theorist might prefer working with linear rather than nonlinear relationships?

9. Give an example of a moderated relationship.

Exercises to Apply Concepts

1. Conduct a thought experiment that describes a relationship between two quantitative variables of your choice. Derive relevant theoretical propositions from this experiment.

2. Conduct a thought experiment that describes a relationship between two qualitative variables of your choice. Derive relevant theoretical propositions from this experiment.

3. Conduct a thought experiment that describes a relationship between a qualitative and a quantitative variable of your choice, using the hypothetical mean approach. Derive relevant theoretical propositions from this experiment.

4. Conduct a thought experiment that describes a relationship between a qualitative and a quantitative variable of your choice, using hypothetical probabilities. Derive relevant theoretical propositions from this experiment.

5. Conduct a thought experiment using a factorial table of hypothetical means with variables of your choice. Derive the full set of theoretical propositions corresponding to main effects, simple main effects, and interaction contrasts.

Appendix 6A

Thought Experiments for a Quantitative Cause and Categorical Effect

A Hypothetical Contingency Table Method

This appendix describes a thought experiment using hypothetical contingency tables to clarify the relationship between a quantitative cause and a categorical effect. It is simpler to apply than the thought experiment described in the main text, though a bit less rigorous. Consider the theoretical proposition that the degree to which a person is liberal affects his or her political party identification. Political party identification is a categorical variable, but liberalness is a quantitative variable that varies from low to high on an underlying continuum. To think through and focus this relationship, we create three categories on the continuum of liberalness: "low," "medium," and "high." We then form a hypothetical contingency table, per the procedures discussed in the main text:

	Democrat	Republican	Independent	Total
High in Liberalness				100
Moderate in Liberalness				100
Low in Liberalness				100

Next we fill in values for each row that we think approximate the real-world state of affairs:

	Democrat	Republican	Independent	Total
High in Liberalness	50	15	35	100
Moderate in Liberalness	25	25	50	100
Low in Liberalness	15	50	35	100

Finally, we translate the variation in numbers within each column into a set of theoretical propositions, using the same principles as discussed for two categorical variables:

> **Proposition 1:** Individuals who are moderate in liberalness are more likely to be Independents than individuals who are low or high in liberalness. Individuals who are low or high in liberalness are equally likely to be Independents.

Proposition 2: Individuals who are low in liberalness are more likely to be Republicans than individuals who are moderate or high in liberalness. Similarly, those who are moderate in liberalness are more likely to be Republican than individuals who are high in liberalness.

Proposition 3: Individuals who are high in liberalness are more likely to be Democrats than individuals who are moderate or low in liberalness. Similarly, those who are moderate in liberalness are more likely to be Democrats than individuals who are low in liberalness.

As always, you should be able to articulate the logic and reasoning underlying your propositions and the entries on the contingency table.

Appendix 6B

Thought Experiments for Moderated Moderation

An interesting but more complex form of moderation analysis is the case where there are two moderator variables, not one, that form what is known in the statistical literature as a three-way interaction. This appendix describes how to conduct a thought experiment for the case of a three-way interaction, often called *moderated moderation*. We give special labels to the four variables involved. As is conventional, X is the focal independent variable, Y is the dependent variable or outcome variable, Z is a *first-order moderator variable*, and Q is a *second-order moderator variable*. The first-order moderator is the variable that is conceptualized as directly moderating the effect of X on Y. The second-order moderator moderates this moderating effect. An example will make this clear.

Let us return to our example where we examined gender differences (X) in relationship quality between mothers and their adolescent child (Y) as moderated by grade (Z) to illustrate moderated moderation. The theory was that gender differences in relationship quality would be stronger for eighth graders than for seventh graders. The following table of hypothetical means was generated:

	Grade 7	Grade 8
Female	6.0	6.0
Male	5.0	4.0

$$IC_1 = (6 - 5) - (6 - 4) = -1$$

The gender difference is indeed larger for eighth graders, as reflected in the interaction contrast value of –1. Suppose that the above table is based solely upon European Americans living in the United States. For Latinos, the table of hypothetical means might look as follows:

	Grade 7	Grade 8
Female	6.0	6.0
Male	6.0	6.0

$$IC_2 = (6 - 6) - (6 - 6) = 1$$

For Latinos, the interaction contrast indicates that the gender difference does not vary as a function of grade. Thus, it appears that the moderating effects of grade depend on the ethnicity of the family. For European American families, grade moderates the effect of gender on relationship quality, whereas this is not true for Latino families. We can formalize this finding by computing the difference between the two interaction contrasts

$$\text{TWIC} = IC_1 - IC_2 = -1 - 0 = -1$$

where TWIC stands for *three-way interaction contrast*, IC_1 is the interaction contrast value for level 1 of Q (European Americans), and IC_2 is the interaction contrast value for level 2 of Q (Latinos). We use the term "three-way interaction contrast" because the comparison maps onto a three-way interaction in traditional analysis of variance.

The important point is that, theoretically, we are suggesting that the qualifying effects of grade on gender differences in relationship satisfaction differ depending on ethnicity. Note that one could, in principle, suggest a third-order moderator variable in which the qualifying effects of Q are moderated by yet another variable, W. For example, the above dynamics may occur for low-socioeconomic-status families but not for high-socioeconomic-status families. This finding would map onto a four-way interaction in traditional analysis of variance models (see Jaccard, 1998, for more details on how to interpret such higher-order interactions).

Part III

FRAMEWORKS FOR THEORY CONSTRUCTION

7

Causal Models

Take away the cause, and the effect ceases.

—MIGUEL CERVANTES (1612)

Causal thinking, and the causal modeling that often goes with it, is probably the most prominent approach to theory construction in the social sciences.[1] In this framework, people (or other entities, e.g., families or organizations) are conceptualized as varying on some construct. Theorists are interested in understanding what *causes* this variation. For example, people differ in how smart they are. The question is, "Why is this?" What *causes* variability in intelligence? People differ in how much money they make. Why is this? What *causes* variability in income? People differ in what they buy, how much they eat, for whom they vote, the organizations they join, and how much of themselves they devote to work. Why is this? What causes this variability? Causal thinking tries to explain variability by identifying its causes.

If something causes variability, then that something also varies. People differ in how smart they are, in part, because they are raised differently by their parents. In this case, variability in intelligence is due to variability in childrearing activities. People differ in how much money they make because they differ in how much education they have. Variability in income is due, in part, to variability in achieved education. Causal analysis involves identifying relationships between variables, with the idea that variation in one variable produces or causes variation in the other variable.

In addition to identifying causes of variables, causal analysis also involves specify-

[1]We use the term *model* instead of *theory* because *model* is typically used in the scientific literature when referring to causal thinking.

ing "effects" of variables. Thus, a theorist might be interested in the consequences of being rich versus poor or the consequences of being stressed versus relaxed. Causal analysis takes many forms, but the essence of causal modeling is the focus on cause–effect relationships.

In this chapter we provide a detailed strategy for building a causal model. We begin by identifying two types of relationships: predictive or associational relationships that are unconcerned with causality, and causal relationships. We then discuss the nature of causality in general as well as the role of the concept of causality in grounded/emergent theories. Six types of relationships are identified that form the core of causal models in the social sciences. These include direct causes, indirect causes, moderated relationships, reciprocal causality, spurious effects, and unanalyzed relationships. Each of these relationships is then elaborated upon in the context of a 10-step approach to constructing a causal model.

TWO TYPES OF RELATIONSHIPS: PREDICTIVE AND CAUSAL

Predictive Relationships

Predictive relationships focus on the question "Is variability in *A* related to variability in *B*?" Note that, at this level, the focus is on mere association; there need be no presumption or implication of causation, only that variations in *A* are related to variations in *B*. If we can identify and verify such a relationship, then we can use our knowledge of variation in *A* to predict variation in *B*, without any need to explain why the association occurs or what causes variability in *B*. For example, one branch of personnel selection is concerned with predicting potential success of job applicants. The goal of this research is to identify variables that will predict this success. In essence, the scientist does not care whether a causal relationship exists between the variables used to predict success and actual success; he or she is only interested in *predicting* job success. In similar fashion, marketing brand managers may be less interested in research that lets them understand what causes consumer purchase behavior and more interested in research that will enable them to *predict* how much their product will be purchased next month. In instances where the focus is on prediction rather than causation, the terminology associated with the two variables are *predictor variable* and *criterion variable*.

An interesting example of a purely predictive orientation is the method used to develop the Minnesota Multiphasic Personality Inventory, also called the MMPI, a widely used test of maladaptive psychological orientations. When the test was constructed, different groups of individuals who had been diagnosed with specific psychological problems were administered a large number of questionnaire items and asked to agree or disagree with each. For example, individuals who had been diagnosed as hypochondriacs indicated their agreement or disagreement with hundreds of items. The same items were completed by a group of "normal" adults, and the responses were then compared for the two groups. Any item that had an agreement pattern that differentiated the two groups became a candidate for inclusion in the final scale, no matter how unusual the item seemed. The result was a subset of 20 or so items to which people with hypochondriasis showed a unique response pattern relative to "normals." If an individual in the general population, when given the MMPI, showed the same response pattern across the items as the hypochondriasis group, then they were declared as likely having hypochondriasis. The MMPI contains some truly bizarre items in terms of content, leading many laypeople who take the test to wonder what exactly is going on. But the test has been carefully developed to have predictive utility, and it often does a reasonable job in correctly diagnosing individuals.

Causal Relationships

Distinct from predictive–associational relationships are causal relationships. These relationships invoke the notion of causality, with the idea that one of the variables in the relationship, *X*, influences the other variable in the relationship, *Y*. The nature of causality has been debated extensively by philosophers of science (e.g., Bunge, 1961; Cartwright, 2007; Frank, 1961; Morgan & Winship, 2007; Pearl, 2000; Rubin, 1974, 1978; Russell, 1931; Shadish, Cook, & Campbell, 2002), and most agree that causality is an elusive concept that is fraught with ambiguities. In fact, the famous philosopher Bertrand Russell (1931) was so flabbergasted by the difficulties with the concept that he suggested the word *causality* be expunged from the English language.

Scientists generally think of causality in terms of *change*. Variable *X* is said to be a cause of *Y* if changes in *X* produce changes in *Y*. Hume (1777/1975) pointed out that it is impossible to ever demonstrate that changes in one variable *produce* changes in another. At best, we can only observe changes in one variable, followed at a later time by changes in another variable. Such coexistent change, he notes, does not necessarily imply causality. For example, an alarm clock going off every morning just before sunrise cannot be said to be the cause of the sun rising, even though the two events are intimately linked.

Russell (1931) argued that causality can be established unambiguously only in a completely isolated system. If one assumes no other variables are present or operating, then changes in *X* that are followed by changes in *Y* are indeed indicative of a causal relation. When contaminating variables are present, however, it is possible for a true causal relationship to exist even though observations show that *X* and *Y* are completely unrelated to each other. Similarly, a causal relationship may not exist, even though *X* and *Y* are found to be related. Having shown this using formal logic, Russell turned to the problem of how one could ever know that one is operating in a completely isolated system, such as in a highly controlled laboratory setting. The only way one can be confident that the system is isolated, he argued, is if changes in *X* unambiguously produce changes in *Y* in that system. But at the same time that we want to assert the existence of an isolated system because changes in *X* produce changes in *Y*, we also want to assert that the reason *X* produces a change in *Y* is because we are operating in an isolated system. Such reasoning, Russell argued, is tautological.

As you might imagine, the underlying issues for conceptualizing causality and for how one establishes causal relationships are complex. They have been debated by extremely bright philosophers of science and scientists for decades. It is beyond the scope of this book to delve into these issues in depth. After reading the relevant literature carefully and giving the matter much thought, it is our belief that, in a strict sense, causality of the type that traditional social scientists seek to infer can rarely, if ever, be demonstrated unequivocally. Strong adherents to experimental methods may take exception to this view, and we respect that. However, we personally find the arguments of Blalock (1964), Bunge (1961), Hume (1777/1975), Russell (1931), and others, taken as a whole, to be convincing (though any single treatise has some flaws).

If causality is so difficult to demonstrate, then why is the concept dominant in social scientific theories? Our answer is that the concept of causality is a heuristic that helps us to think about our environment, organize our thoughts, predict future events, and even change future events. By thinking in causal terms, we are able to identify systematic relationships between variables and manipulate those variables so as to produce changes in phenomena that are scientifically or socially desirable to change. For example, through causal thinking, we might derive a certain type of therapy designed to reduce the severity of migraine headaches. Although technically we may never be able to establish a causal link between the therapy and the reduction of headache severity, the fact of the matter is that applying the treatment accomplishes the socially desirable goal of headache reduction. If causal thinking aids in the development of such programs, then the causal theory that leads to this state of affairs is useful.

Stated another way, we may never be able to unambiguously demonstrate causality between variables, but we certainly can have differing degrees of confidence that a causal relationship (of the form that "changes in *X* produce changes in *Y*") exists between variables. Scientific research is conducted to establish strong, moderate, or weak levels of confidence in theoretical statements that propose causal relationships.

There are some common features of the construct of causality upon which most social scientists agree. First, as noted, if *X* causes *Y*, then changes in *X* are thought to produce changes in *Y* (however, see Sowa, 2000, and Lewis, 2000, for alternative conceptualizations). Second, a cause always must precede an effect in time. Third, the time that it takes for a change in *X* to produce a change in *Y* can vary, ranging from virtually instantaneous change to years, decades, centuries, or millennia. Fourth, the nature and/or strength of the effect of *X* on *Y* can vary depending on context. *X* may influence *Y* in one context but not another context. Finally, cause and effect must be in some form of spatial contact or must be connected by a chain of intermediate events. We return to each of these points in later sections of this chapter.

Not all scientific theories rely on the concept of causality. In fact, certain areas of physics did not progress until the notion of causality was deemphasized or virtually abandoned (see Sowa, 2000). Nevertheless, causality remains the dominant system of thought in the social sciences.

CAUSALITY AND GROUNDED/EMERGENT THEORY

Causal explanation has been the subject of controversy among some grounded/emergent theorists (Maxwell, 2004), especially as it is typically treated in the philosophy of science. There is a vocal group of grounded theorists who have challenged traditional views of causality and who offer perspectives that are more consistent with qualitative methods and process-oriented explanations. Among the alternative frameworks are causal realism (Salmon, 1984, 1989, 1998), constructive empiricism (van Fraassen, 1980, 1989), and ordinary language philosophy (Achinstein, 1983), to name a few. One of the more popular alternatives, causal realism, argues for a real though not "objectively" knowable world. Causal realism holds that phenomena within the objective world are so intertwined and so dependent on one another in such complex ways that simple variable-centered notions of causal regularities are inadequate. There are several variants of causal realism, but we do not digress into these here. Our focus in this chapter is on the more dominant variable-centered approaches to causal explanation that conceptualize causality as described earlier. We consider the other approaches to explanation in Chapter 10. Even if theorists are committed to process-oriented perspectives or other explanatory frameworks, we believe it is helpful to temporarily "think outside the box" by thinking about direct relationships, indirect relationships, moderated relations, spurious relationships, reciprocal causes, and feedback loops. Doing so might provide fresh insights into eventually framing theory in the particular way one wants to frame it.

TYPES OF CAUSAL RELATIONSHIPS

When two variables have a causal relationship, the presumed cause is sometimes called the *independent variable* or the *determinant*. The presumed effect is called the *dependent variable* or *outcome variable*. Causal models have one or more of six types of "causal" relationships in them. The six relationships capture the universe of relationship types used in causal modeling. In this section we briefly characterize these six types of relationships to provide you with an overview of them. Then, we delve into each type of relationship in detail. The six relationships—(1) direct causal, (2) indirect causal, (3) spurious, (4) moderated causal, (5) bidirectional causal, and (6) unanalyzed—are shown in Figure 7.1. In this figure a variable is indicated by a box and a causal influence is represented by a straight arrow emanating from the cause and pointing to the effect. We discuss the curved arrow later. This type of figure is called a *path diagram*.

A *direct causal relationship* is one in which a given cause is assumed to have a direct causal impact on some outcome variable. For example, frustration is assumed to influence aggression. As another example, the quality of the relationship between a mother and her adolescent child is assumed to influence whether the child uses drugs, with poor relationships being associated with higher levels of drug use. Figure 7.2a illustrates this latter relationship.

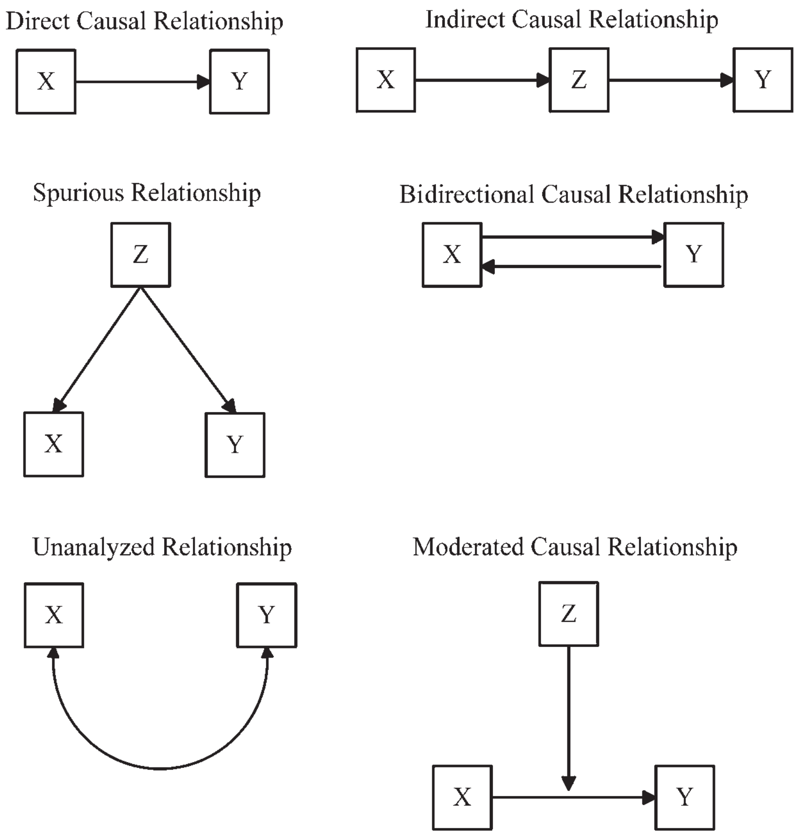


FIGURE 7.1. Relationships in causal models.

An *indirect causal relationship* is one in which a variable influences another variable indirectly through its impact on an intermediary variable (see Figure 7.1). For example, failing to accomplish a goal may lead to frustration, which, in turn, causes someone to aggress against another. In this case, failure to obtain a goal is an indirect cause of aggression. It only influences aggression through its impact on frustration. Frustration is formally called a *mediating variable*, or more informally, a *mediator*, because other variables "work through" it to influence the outcome. Indirect relationships are sometimes called mediated relationships.

Figure 7.2b illustrates an indirect relationship between the quality of the relationship between a mother and her child and adolescent drug use. The quality of the relationship is assumed to impact how much the adolescent orients toward working hard in school, with better relationships leading to working harder. Students who work hard in school, in turn, are assumed to be less likely to use drugs. Figures 7.2a and 7.2b illustrate an important point: What is a direct relationship in one theory can be an indirect relationship in another theory. We comment more on this point later.

A *spurious relationship* is one in which two variables are related because they share a common cause, but not because either causes the other (see Figure 7.1). As an example, if we select a random sample of all people in the United States and calculate the correlation between shoe size and verbal ability, we would find a moderate relationship between the two variables: People with bigger feet have more verbal ability. Does this mean that a causal relationship exists between these variables? Of course not. The reason they are correlated is because they share a common cause: age. A random sample of people in the United States will include large numbers of children. When children are young, they have small feet and they can't talk very well. As they get older, their feet grow and their verbal ability increases. The common cause of age produces a correlation between shoe size and verbal ability, but it is completely spurious.

A *moderated causal relationship*, like spurious and indirect relationships, involves at least three variables (see Figure 7.1). In this case, the causal relationship between two variables, X and Y, differs depending on the value of a third variable, Z. For example, it might be found that a given type of psychotherapy (X) is effective for reducing headaches (Y) for males but not for females. In this case, the causal relationship between psychotherapy and headache reduction is moderated by gender. When gender has the value "male," X impacts Y. However, when gender has the value "female," X does not impact Y. Gender is called a *moderator variable* because the relationship between the presence or absence of psychotherapy (X) and headache reduction (Y) changes as a function of (or is "moderated by") the levels of gender.

A *bidirectional or reciprocal causal relationship* exists when two variables are conceptualized as influencing each other (see Figure 7.1). For example, in the area of reproductive health, a theorist might posit a bidirectional influence between a woman's belief that the rhythm method is effective at preventing pregnancy (X) and her attitude toward the rhythm method (Y). A woman may have a positive attitude toward the rhythm method because she believes it is effective. Simultaneously, she may believe it is effective, in part, because she has a positive attitude toward it, via a mechanism that involves rationalization of behavior.

The final type of relationship is an *unanalyzed relationship*. In Figure 7.1 the two variables for this type of relationship are connected by a double-headed curved arrow. This arrow signifies that the two variables are correlated, but that the theorist is not going to specify why they are correlated. The correlation may be spurious or it may be

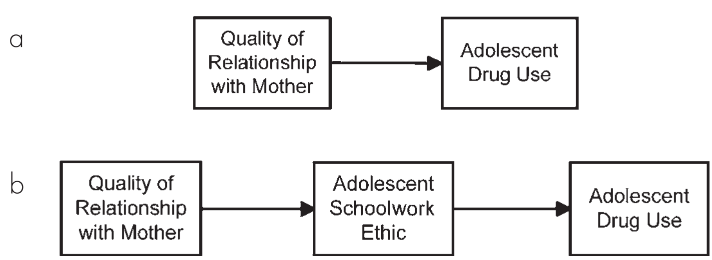


FIGURE 7.2. Examples of direct and indirect relationships. (a) Direct relationship; (b) indirect relationship.

due to a causal connection of some kind. The theorist wants to recognize the correlation between the variables, but trying to explain it is beyond the scope of the theoretical effort. Examples of this approach are illustrated later in this chapter.

Most causal models have more than one of these six types of relationships in them. We provide an example of a multivariate causal model in Figure 7.3. In this model there are several direct relationships. How hard an adolescent works in school is assumed to be a direct cause of drug use. The quality of the relationship between the mother and child is assumed to be a direct cause of how hard the adolescent works in school. The quality of the relationship between the mother and child has an indirect causal relationship with drug use that is mediated by how hard the child works in school. The amount of time that a mother spends with her child is assumed to have a direct influence on the quality of the relationship between the parent and child. Gender of the adolescent is assumed to have a direct impact on the amount of time that a mother spends with her child, with mothers spending more time with girls than boys. Gender also has a direct influence on the quality of the relationship between mothers and their children, with mothers having better relationships with girls than boys. Note that because gender influences both the amount of time spent with the child and the quality of the relationship between mother and child, it is a common cause for these variables. Hence, some of the association between time spent together and relationship quality is spurious. However, the straight causal arrow between the time spent together and relationship quality indicates that the theorist believes some of the association is not spurious. Rather, there also is some true causal influence operating. In this case, the association between the two variables is thought to have two components: (1) a causal component and (2) a spurious component.

Note also in this model that the amount of time a mother spends with her adolescent is an indirect cause of drug use. The indirect effect works through two sequential mediators: (1) the quality of the relationship with the child and, in turn, (2) how hard the child works in school. There are several other more distal indirect relationships in this model. Try to identify them.

There are no moderated relationships, nor are there any unanalyzed relationships

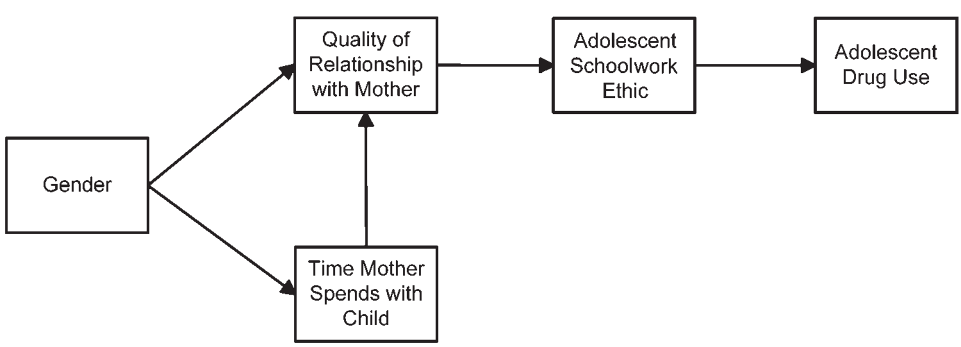


FIGURE 7.3. Multivariate causal model.

in this model. It is not necessary for a theory to contain all six types of relationships we described earlier.

A common distinction in causal theories is between exogenous and endogenous variables. An *endogenous variable* has at least one causal arrow pointing to it. An *exogenous variable* does not have a causal arrow pointing to it. In Figure 7.3, for example, gender is an exogenous variable, and all other variables in the system are endogenous variables. We occasionally use this terminology.

CONSTRUCTING THEORIES WITH CAUSAL RELATIONSHIPS

We now discuss a process for constructing causal models. We draw on the heuristics discussed in Chapter 4 while explicating in more depth the six types of relationships in Figure 7.1. It is not possible to convey this material in a straightforward, linear fashion, so be prepared for digressions. The approach we describe is not the only way to develop a causal theory. It represents one approach that we find useful, but we often deviate from it ourselves in our own theory construction efforts.

IDENTIFYING OUTCOME VARIABLES

The approach we describe involves first identifying an outcome variable and then specifying some causes of that variable. Thus, the first step is choosing an outcome variable that you want to explain. Perhaps you are interested in understanding why some people are Republicans but other people are Democrats. Or you might be interested in understanding why some people become alcoholics but others do not. In Chapter 4 we discussed strategies for choosing outcome variables. Choose one now.

Some researchers approach theory construction using a reverse process; that is, they specify a variable of interest and then ask what are the consequences or "effects" of it. For example, a theorist might be interested in self-esteem and what the consequences are of having low versus high self-esteem. Or a theorist might be interested in poverty and want to explore the effects of poverty on people's lives. There is more than one way to go about theory construction. If you prefer the latter approach, then the concepts and strategies we discuss below are still applicable, but will have to be adapted somewhat. We specify when such adaptations are required.

Some researchers decide to build a theory of the effects of an intervention on an outcome. For example, an educational researcher might develop a program to improve reading skills and plan to build a theory around its effects on reading. He or she plans to conduct a study in which an experimental group receives the intervention and a control group receives the standard reading curriculum. In this case, you already have a designated "cause" or independent variable (the intervention group vs. the control group), and you also have identified an outcome variable or "effect," reading ability.

IDENTIFYING DIRECT CAUSES

We start the theory construction process by specifying two or three variables that are direct causes of the outcome variable. We do not specify more than two or three direct causes at this initial step because we ultimately will subject each direct cause to considerable elaboration. The theory might become overwhelming at later stages if we work with too many variables initially. Additional direct causes always can be added at a later point.

Use the heuristics and strategies discussed in Chapter 4 to identify your initial set of direct causes, and use the strategies discussed in Chapter 5 to ensure that the concepts with which you are working are appropriately focused. You also can apply the thought experiments described in Chapter 6 to clarify the relationships. When specifying a direct cause, remember that your goal is to explain why there is variation in the outcome variable you have chosen. If the outcome variable is popularity, for example, then you want to know what causes people to differ in their popularity. What makes some people more popular than others? What factors influence popularity? If the outcome variable is teacher apathy in schools, then you want to know why some teachers are apathetic and other teachers are not. When specifying your direct causes, keep this focus in mind.

If you adopt the strategy of choosing an initial variable but want to treat it as a cause rather than an effect, then identify two or three variables that the variable is thought to impact. For example, if the primary variable of interest is Latino acculturation to U.S. culture, you might use drug use as one "effect" (under the tentative presupposition that increased acculturation increases drug use) and performance in school as another "effect" (under the tentative presupposition that increased acculturation increases performance in school). Whichever approach is used, you should now have a theory with two to three direct effects in it. Be sure that you can articulate the logic underlying each direct effect in your model.

Finally, if you are building a theory of the effects of an intervention, you already specified (from the previous section) a direct effect of the intervention (intervention vs. control) on the outcome (e.g., reading ability). In this case, we will work with this single direct cause.

To make the next tasks manageable, we recommend that you draw your theory using a path diagram, which is a pictorial representation of a theory, using boxes and arrows as in Figures 7.1–7.3. By the end of this chapter, your path diagram will be complex, but at this stage, it should be simple, consisting of a few direct causes, in the spirit of a direct effect in Figure 7.1. As you complete each step that follows, continually update your path diagram. The steps will add complexity to your theory, and you will need to use the path diagram as an aid to avoid being overwhelmed by what is to come.

INDIRECT CAUSAL RELATIONSHIPS

Turning Direct Relationships into Indirect Relationships

Once you have identified a few direct causes, the next step is to try to turn the direct causes into indirect causes. That is, we identify variables that mediate the direct relationships and then insert these variables into the theoretical system. For example, suppose that our outcome variable is drug use in adolescents and one of the direct causes we identified is the quality of the relationship with the mother of the adolescent. We expect that adolescents with better relationships with their mothers will be less likely to use drugs. If we ask ourselves the question "Why do you think that quality of the relationship impacts drug use?" we might answer that adolescents who have a good relationship with their mothers work harder in school in order to please them, and this increased focus on school is why adolescents are less likely to use drugs. Contained in this answer is a mediator variable, namely, the increased focus on school. What was a direct relationship now can be turned into an indirect relationship: Quality of the relationship impacts the adolescent's focus on school, which, in turn, impacts drug use. We refer to this strategy as the *why heuristic*.

We can take this a step further and attempt to turn the newly established direct relationship between "school focus" and "drug use" into an indirect relationship. We ask ourselves "Why do you think working hard and focusing on school impacts drug use?" We might answer "because then adolescents have less time for after-school activities that expose them to drug use." We now have a new mediator, namely, avoidance of risk situations. It can be used to turn the direct relationship between "school focus" and "drug use" into an indirect relationship using "avoidance of risk situations" as a mediator. This new variable is somewhat vague, and we need to apply the focusing strategies discussed in Chapter 5 to clarify it. But that is not the point here. The main idea is that you can expand a theoretical framework that has direct causes by turning direct causal relationships into indirect causal relationships through the specification of mediators. You continue to do this until you reach a point where you just don't want to further explicate mediators. That is, you reach a point where you want to close this aspect of the theoretical system and move onto other features of the theory.

In sum, to turn a direct causal relationship into an indirect causal relationship, ask yourself the question "Why is it that *X* influences (i.e., reduces or increases) *Y*?" As you articulate your answer to this question (substituting the actual variables names for *X* and *Y*), therein will lie a potential mediator variable. Why is it that higher levels of education lead to higher levels of income? Your answer to this question is a potential mediator. Why is it that boys drink more alcohol than girls? Your answer to this question is a potential mediator.

Partial versus Complete Mediation

Once you have specified a mediator and added it to your path diagram, you are confronted with a new issue. Examine Figure 7.4a, which shows an indirect relationship where the impact of *X* on *Y* is mediated by *Z*. According to this model, the *only* way in which *X* influences *Y* is through *Z*. Stated another way, *Z* completely mediates any impact *X* has on *Y*. Therefore, *Z* is a *complete mediator*.

But another possibility exists. Maybe *Z* only partially mediates the effects of *X* on *Y*. Perhaps in addition to the mediated effects of *X* on *Y* through *Z*, *X* also has an independent effect on *Y* that can't be accounted for by *Z*. This scenario is illustrated in Figure

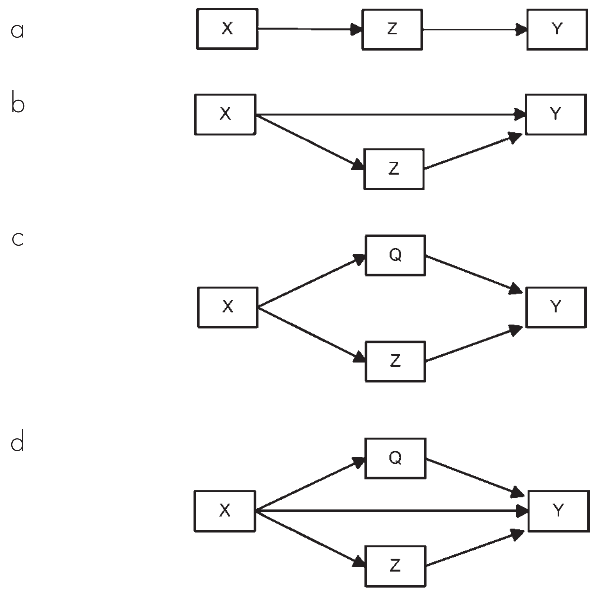


FIGURE 7.4. Complete and partial mediation. (a) Complete mediation; (b) partial mediation; (c) complete mediation with two mediators; (d) partial mediation with two mediators.

7.4b. In this case, *Z* is said to be a *partial mediator* of the effect of *X* on *Y*. As an example, in Figure 7.2b, the quality of the relationship with the mother impacts the adolescent's work ethic in school, which, in turn, influences the adolescent's drug use. Perhaps in addition to these effects, the quality of the relationship with the mother has an independent effect on drug use, over and above its effect through the adolescent work ethic. If so, this represents partial mediation: The adolescent's school work ethic mediates some of the impact of quality of the maternal relationship on drug use—but not all of it.

In any causal system, once you introduce a mediator, you must next decide if the mediator is a complete or partial mediator. After inserting the mediators into your path diagram, you must further adjust the theory either by drawing arrows to represent partial mediation or excluding arrows to reflect complete mediation, as per Figure 7.4.

What if you are not sure which to specify, complete or partial mediation? Here is the approach we use in such cases. For partial mediation, you are essentially stating that there is some mechanism other than *Z* by which *X* influences *Y*. What is that other mechanism? If you can articulate it, then partial mediation is called for; if you cannot articulate it, then complete mediation is the answer. In essence, we take the direct effect between *X* and *Y* in Figure 7.4b and try turning it into an indirect effect by identifying a second mediator, *Q*. This is illustrated in Figure 7.4c. If we can identify *Q*, then partial mediation in the model is called for; if we can't think of *Q*, then complete mediation is the answer. Continuing with our drug use example, we might conjecture that in addition to adolescents' schoolwork ethic, the quality of the mother–adolescent relationship also impacts how much adolescents are willing to allow their mothers to keep track of their activities on weekends: If the relationship between the mother and adolescent is poor, then the adolescent will resist attempts by the mother to keep track of or monitor him or her. If the relationship is good, then the adolescent may not resist as much. Thus, the quality of the mother–adolescent relationship (X) impacts not only the adolescent's schoolwork ethic (Z) but also parental monitoring (Q), and both of these variables (X and Q) are thought to impact adolescent drug use. In this case, we are justified in hypothesizing partial mediation, as per Figure 7.4b, because we are able to specify a reasonable mechanism for it.

If we specify Q, then why not just incorporate it into the theory? Of course, we could very well do this, but then the issue becomes whether the two mediators, Z and Q, considered together, are complete or partial mediators of the causal effect of X on Y. That is, perhaps now the model should appear as in Figure 7.4d instead of Figure 7.4c. To add the direct path from X to Y over and above Q and Z, we would need to be able to articulate yet a third mediator. At some point, you decide to close out the system and just let a direct path between X and Y stand so as to reflect partial mediation without formally bringing additional mediators into the model. If pressed, you could articulate one, but you simply do not want to complicate the theory further.

An Alternative Strategy for Turning Direct Effects into Indirect Effects

There is another way to bring indirect causal relationships into the theory. Pick one of your direct causes, X, and now treat it as an outcome variable. Then use the heuristics discussed in Chapter 4 to identify causes of this cause. Identify a few such causes and add them to your path diagram. You will now have indirect relationships between these new causes and the original outcome variable that are mediated by your initial direct cause. The variable that originally took the role of a direct cause now takes on the additional role of mediator. We call this strategy a *cause of a cause heuristic*.

Figure 7.5 illustrates this dynamic for the drug use example. The initial direct cause was the quality of the relationship between the mother and the adolescent (see Figure 7.5a). We then treat the quality of the relationship as an outcome variable and ask what factors influence it. We might conjecture that the gender of the child impacts the quality of the relationship and then add this to the theory (see Figure 7.5b). Now the original direct cause is a mediator. Note that at any point in the process, we can try to turn a direct cause into an indirect cause using our first strategy of answering the question of "why." For example, "Why is it that the gender of the adolescent influences the quality of the relationship between the mother and the adolescent?" Our answer to this question might be "because mothers spend more time with girls than boys," which yields a mediator. This causal dynamic is illustrated in Figure 7.5c, which further augments the theoretical system. Also, any time we create a mediator, we also must make decisions about complete or partial mediation. In Figure 7.5c, we have assumed complete mediation.

Of course, there will be some causes, such as gender or race, where it does not make

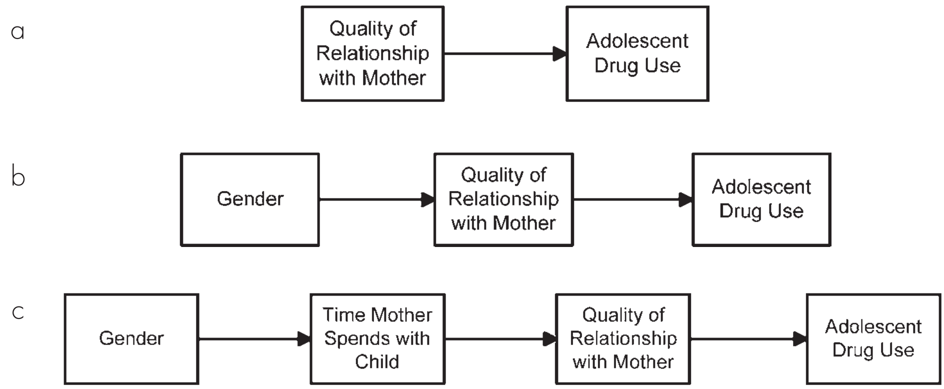


FIGURE 7.5. Example of making the cause an outcome. (a) Direct relationship; (b) quality of relationship becomes an outcome; (c) inserting a mediator.

sense to treat them as an outcome variable and where this strategy is inappropriate. This will also be true when the "cause" is an intervention.

Summary of Mediation

In sum, there are two heuristics for creating indirect effects in your model. First, the *why* heuristic involves focusing on a direct causal relationship between X and Y and asking "Why does X influence Y?" The answer to this question contains the mediator. Second, the cause of a cause heuristic treats one of your direct causes as an outcome and identifies causes of it. Once you add a mediator, you must decide about partial or complete mediation with respect to it. Complete mediation is called for if you are unable to articulate any mechanism other than Z by which X impacts Y. Partial mediation is called for if you can specify such a mechanism. Apply these two heuristics to the direct relationships in your model and draw the mediated relationships you have identified into your path diagram. Add partial mediation causal arrows (direct effects), if appropriate.

MODERATED CAUSAL RELATIONSHIPS

The next step in the theory construction process is to consider the addition of moderated causal relationships. As discussed in Chapter 6, a moderated causal relationship involves three variables, a cause (X), an effect (Y), and a moderator variable (Z). The essence of a moderated relationship is that the strength or nature of the effect of X on Y varies as a function of Z. Examine every direct relationship that is in your theory as drawn in your path diagram. For each relationship, ask yourself, "Are there some circumstances where the impact of X on Y will be stronger than in other circumstances?" If the answer is "yes," try to articulate those circumstances and identify their defining characteristics. As you do so, you will be describing your moderator variable. In other words, carefully examine the circumstances you generate and try to abstract a variable that captures or represents them. Use the methods in Chapter 5 to focus this variable and then use the methods in Chapter 6 to clarify the moderated relationship.

As another approach, for every direct cause in your theory, ask yourself if the impact of *X* on *Y* will be stronger for some individuals than other individuals. If the answer is "yes," try to describe the defining characteristics of the individuals for whom it will be stronger. As you do so, you will be describing your moderator variable.

Of course, you may not want to pursue this strategy for every direct cause in your theory, but the potential for doing so exists. Draw the moderated relationships you have identified into your path diagram. Your theory now should include direct causal relationships, indirect causal relationships with either partial or complete mediation, and moderated relationships.

Mediated Moderation

Next, you should consider the possibility of adding a mediated moderator relationship. This type of relationship combines an indirect and moderated relationship; it is diagrammed in Figure 7.6a. Note that *Z* is a traditional moderator variable that impacts the strength of the effect of *X* on *Y*. However, we have inserted a mediator of the moderating effect, *Q*, into the model. For example, suppose we find that gender (*Z*) moderates the impact of a multisession psychotherapeutic treatment (*X*) on depression (*Y*), such that the treatment is more effective for females than males. Asking "*Why* is the treatment more effective for females than males?", we might conjecture that females are more likely than males to attend all of the therapy sessions. By answering the question *why* the moderation occurs, we identify a mediator (*Q*) of that moderation, which in this case is session attendance. In other words, we use the *why* heuristic to generate a mediator. Draw any mediated moderators into your model that you care to add now.

As before, any time you add a mediator, there is the possibility of complete or partial mediation. The same is true for mediated moderation. It may be that the mediator accounts for only some of the moderating effects of *Z* on the effect of *X* on *Y*. Figure 7.6b illustrates a case of partial mediated moderation. Now determine for your theory if you want to assume complete or partial mediated moderation.

Moderated Mediation

Next, consider the possibility of adding a moderated mediated relationship. Moderated mediation occurs when the strength of a path signifying mediation varies as a function of some variable, *Q* (see Figure 7.6c). For example, for females, *Z* might be a complete mediator of the effect of *X* on *Y*, whereas for males, *Z* may be only a partial mediator of the effect of *X* on *Y*. In this case, the mediational properties of *Z* depend on the value of the moderator variable, *Q*. Moderation can occur at any or all of the paths involved in mediation. In Figure 7.6c, moderation operates at the level of two of the paths involved

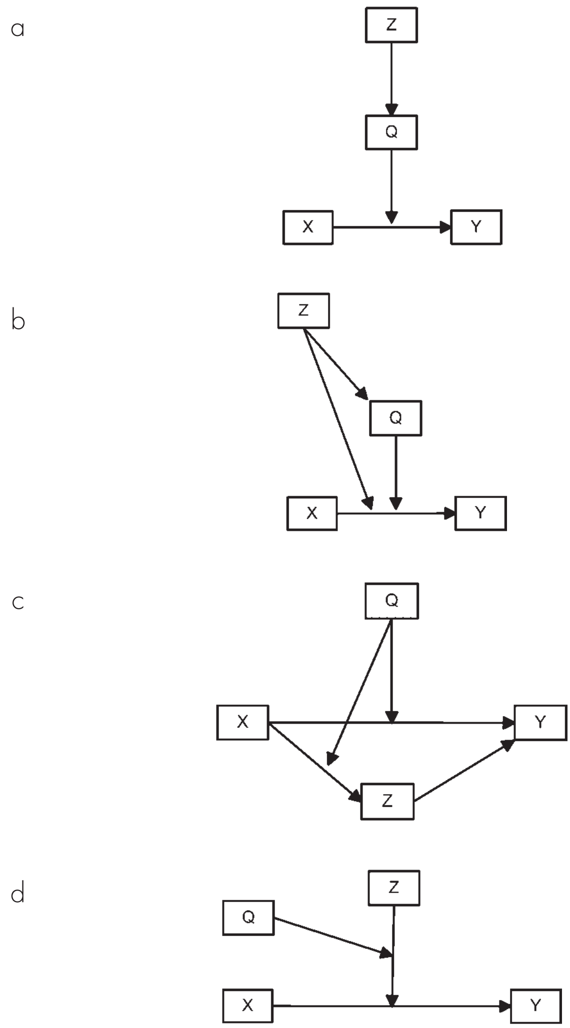


FIGURE 7.6. Mediated moderation, moderated mediation, and moderated moderation. (a) Mediated moderation; (b) partial mediated moderation; (c) moderated mediation; (d) moderated moderation.

in mediation, but not the third path. Use the above heuristics to identify relevant moderator variables. Be sure that you can articulate the logic underlying the modifications you make to the theory.

Moderated Moderation

A final possibility to consider is a moderator of a moderated relationship. This is diagrammed in Figure 7.6d, where *Q* moderates the moderating qualities of *Z*. We give special labels to *X*, *Z*, and *Q*. *Y* is the outcome variable, X is the focal independent variable, Z is a *first-order moderator variable*, and Q is a *second-order moderator variable*. The first-order moderator is conceptualized as directly moderating the effect of X on Y. The second-order moderator moderates this moderating effect. See Appendix 6B of Chapter 6 for details and an example. Consider if you want to include moderated moderation in your theory.

Summary of Moderated Relationships

Moderated relationships can be incorporated into a theory by asking questions about whether the effect of X on Y will be equally strong in all circumstances or whether the effect of X on Y will be equally strong for all individuals—that is, using the *stronger-than heuristic*. As noted in Chapter 6, this heuristic involves asking if the effect of X on Y will be stronger in some circumstances than in others or stronger for some individuals than for others. The possibility of a moderated relationship can be considered for all direct causes in the theory as well as for mediated relationships in the form of moderated mediation. If you add a moderator variable, then you should consider the possibility of adding mediated moderation, either partial or complete. You also can consider moderated moderation. Draw these dynamics into your path diagram, as appropriate. Articulate the logic of each path that you add and focus the concepts and relationships using the methods described in Chapters 5 and 6.

RECIPROCAL OR BIDIRECTIONAL CAUSALITY

There Is No Such Thing as Simultaneous Reciprocal Causality

Reciprocal or bidirectional causal relationships occur when a variable, X, influences another variable, Y, and Y also influences X (see Figure 7.1). Strictly speaking, there can never be simultaneous reciprocal causation because there always must be a time interval, no matter how infinitesimally small, between the cause and the effect that follows from that cause. If we observed the causal dynamics within the appropriate time frames, the true dynamic underlying a reciprocal causal relationship would appear as follows

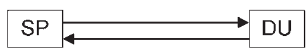
$$X_{t1} \rightarrow Y_{t2} \rightarrow X_{t3} \rightarrow Y_{t4}$$

where X_{t1} is variable X at time 1, Y_{t2} is variable Y at time 2, X_{t3} is variable X at time 3, and Y_{t4} is variable Y at time 4. It is only when we are unable to capture the appropriate time intervals, and must instead work with coarser time intervals, that the dynamic of the reciprocal causal relationship, as illustrated in Figure 7.1, applies. Essentially, by working with coarser time units, the more fine-grained temporal causal dynamics already have played themselves out (which is known in the causal modeling literature as the *equilibrium assumption*). Conceptually, we are working with variables that now reflect the alternating causal dynamics that operated across the more fine-grained time interval in the past. There is nothing inherently wrong with theorizing at the level of coarser time units, as long as we appreciate the underlying logic.

As an example, consider performance in school as measured by grade-point average and drug use by adolescents. It is likely that performing poorly in school puts adolescents at risk for drug use, as their interests drift away from doing well in school. At the same time, school performance is probably adversely affected by drug use, interfering with students' ability to complete their homework as well as adversely affecting their concentration on tests. A causal chain that describes this dynamic is

$$SP_{t1} \rightarrow DU_{t2} \rightarrow SP_{t3} \rightarrow DU_{t4}$$

where *SP* represents school performance at time *t*, *DU* represents drug use at time *t*, and the numerical subscript attached to *t* represents more distant time intervals as the numbers increase in value. If one is unable to assess these processes at the finer-grained time intervals where the causal dynamics are operating, and if these processes have already played themselves out when the assessments of drug use and school performance are made, then the resulting causal representation that captures what has transpired is this:



This representation reflects a summary of the sequential dynamics that have already transpired at a given cross-section in time.

As a next step in the theory construction process, consider introducing reciprocal causality into the system. This should not be done in too cavalier a fashion in the interest of parsimony and the difficulties that reciprocal causation can create for empirical tests of the theory. But if you believe a reciprocal relationship is called for and that it is theoretically important, then include it.

Feedback Loops: Adding Mediators to Reciprocal Causation

Theories sometimes include feedback loops, an example of which is shown in Figure 7.7a. Variable *X* influences variable *Z*, which, in turn, influences variable *Y*, which, in turn, "feed backs" to influence variable *X*. For example, how satisfied supervisors are with their workers (*X*) may impact how satisfied employees are with their jobs (*Z*)—that is, workers like their job better if their boss is happy with them. Employee job satisfaction may, in turn, impact the productivity of employees (*Y*), with more satisfied workers being more productive. The productivity of employees, in turn, may "feed back" and impact how satisfied supervisors are with their workers (*X*). Such feedback loops are merely a reciprocal causal relationship with a mediator variable inserted into the causal chain. This is evident if we redraw Figure 7.7a in the format of Figure 7.7b.

Add any mediators to your reciprocal causal relationships using the *why heuristic* we discussed earlier. You can add mediators to either one or both causal paths in the reciprocal relationship. We illustrate the latter case in Figure 7.7c.

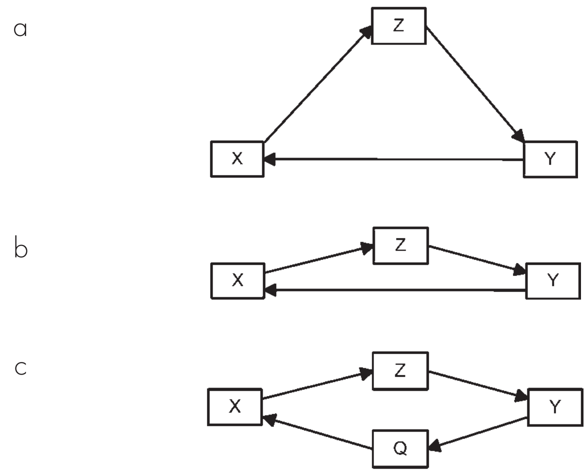


FIGURE 7.7. Feedback loops as reciprocal causation. (a) Traditional feedback path diagram; (b) redrawn feedback loop; (c) feedback loop with two mediators.

Moderated Reciprocal Causation

Reciprocal dynamics may operate in some situations but not others or for some individuals but not others. This point suggests that moderator variables can be added to one or both of the reciprocal causal paths. Figure 7.8a illustrates the case of a moderator variable associated with one causal path in the reciprocal causal relationship, and Figure 7.8b illustrates the case of two moderator variables, one associated with each causal path. Figure 7.8c illustrates the case of a single moderator variable associated with both causal paths. Of course, you can add multiple moderators, mediated moderators, or moderated moderators when introducing moderator variables at this juncture. Your "theoretical toolbox" allows you a great many possibilities. Moderators can be identified using the *stronger-than* heuristic. Make additions to your path diagram accordingly and articulate the underlying logic for these modifications.

SPURIOUS RELATIONSHIPS

In the theory construction process we are elaborating, we typically do not set out to create spurious relationships within the theory. Rather, spurious relationships naturally emerge as we work through the other facets of theory construction. The next steps we recommend often create spurious effects within a theory. Before describing these steps, we must emphasize that spurious effects are not inherently bad, nor are they something to be avoided. In empirical research that tests theories, critics often question a theoretical test by claiming that an observed relationship in the data used to assert a direct causal relationship may, alternatively, represent a spurious relationship. It is one thing to criticize a scientist for conducting a flawed empirical test of a proposed causal link,

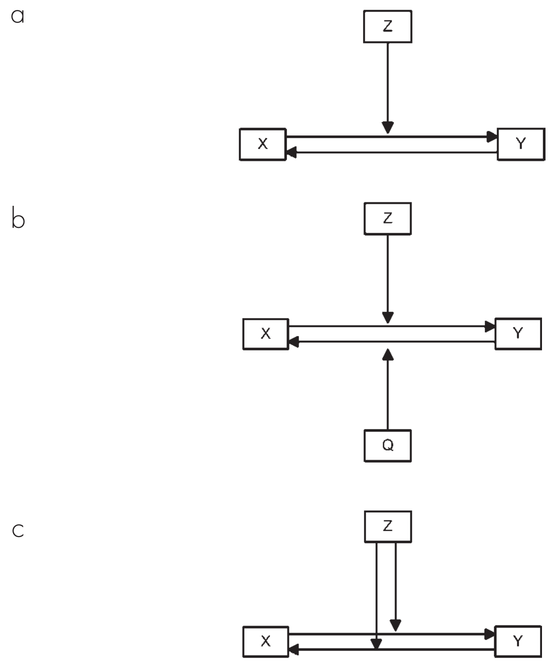


FIGURE 7.8. Moderated reciprocal causation. (a) One-moderator variable model; (b) two-moderator variable model; (c) one moderator variable, two moderated relationships.

but this is not the same as recognizing that many phenomena have common causes in the real world. For example, a fear of contracting AIDS might simultaneously influence one's use of condoms, the number of sexual partners one has, and how frequently one engages in sex. The latter three variables should exhibit some correlation with each other because they share the common cause of fear of AIDS. These correlations are not artifactual. They reflect the operation of a meaningful common cause in real life, and social scientists should embrace them conceptually.

We now discuss three additional steps to consider when building your theory, each of which can create spurious relationships in the theory.

Adding Additional Outcomes

First, consider adding more outcome variables to your theory. Recall that the first step in the theory construction process was to identify a single outcome variable that you were interested in explaining. Consider other such outcome variables at this time, variables that are conceptually related to your initial outcome variable. For example, if your initial outcome variable was use of condoms, perhaps you might add the number of sexual partners and frequency of sexual intercourse as outcomes. Maybe it would be of interest to map the effects of the direct causes you initially specified for condom use onto these variables as well. On the other hand, you may choose not to add other outcomes, deciding that the system is appropriately focused on the one outcome you initially chose. If you add new outcome variables, then you must specify how *all* of the variables currently in your theory are related to them by adding appropriate causal paths.

Adding Effects of Effects

Another strategy for expanding your theory is to think of the original outcome variable that you started with and treat it as a cause of some new variable. In other words, make your effect a cause. What variables might your original outcome variable influence? Add these variables to the theory and draw causal arrows from your outcome variable to them. Note that in doing so, you have turned your original outcome variable into a mediator variable that mediates the effects of your original direct causes on your new outcome variables. With the "new" mediator, you must take a position on partial or complete mediation. Also consider the possibility of adding a moderator variable to the new direct effect as well as mediated moderation, moderated mediation, and/or moderated moderation to it. As always, be sure you can articulate the underlying logic of every path in your model. Update your path diagram accordingly.

Specifying Causal Relationships between Existing Variables

Finally, for all of the variables in your theory at this time, map out the causal pathways between them. As an example, Figure 7.9a represents how your theory may have looked

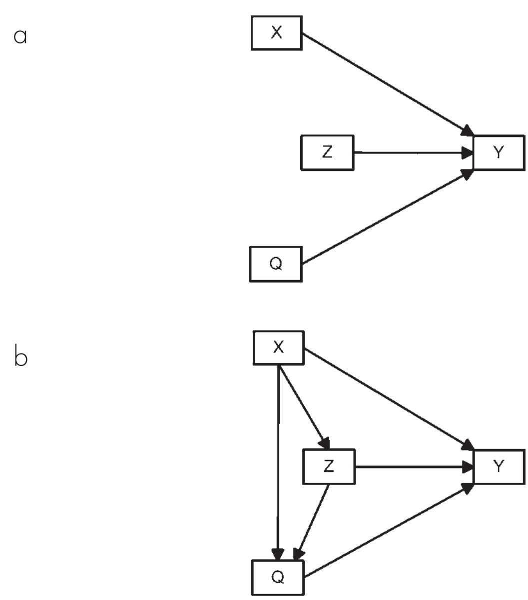


FIGURE 7.9. Mapping causal relationships among all variables. (a) Original specification; (b) mapped specification.

after the step of identifying an outcome variable, Y, and then adding a few direct causes, X, Z, and Q. At this stage, you had made no statements about the causal relationships between X, Z, and Q. Now is the time to consider them. Could X influence Q or Z? Could Z influence Q or X? Could Q influence X or Z? Figure 7.9b shows one example of causal relationships you might impose on the existing variables. As you create new direct or indirect effects, consider elaborating them using all of the tools we have described before (e.g., mediation, partial mediation, moderators, moderated mediation, mediated moderation, and moderated moderation).

Summary of Additional Steps That May Create Spuriousness

To further expand your theory, you will want to consider the strategies of (1) adding additional outcome variables, (2) treating your outcomes as causes and specifying their effects, and (3) mapping out the causal relations among all variables within the theory. The results of these steps may create some spuriousness in the system, but the spuriousness should be of theoretical interest.

UNANALYZED RELATIONSHIPS

In causal models it is typically the case that one is uninterested in causal relationships between the exogenous variables. Causal relationships might exist between them, but you must close out the theoretical system at some point, and elaborating these casual relations is of secondary importance. Hence, you choose to ignore these causal dynamics, but you need to recognize that the exogenous variables are correlated. For this reason, it is traditional to create unanalyzed relationships between all the exogenous variables in a causal model, unless there is a strong theoretical reason for saying there is a zero correlation between them. Figure 7.10a shows a model without the unanalyzed relationships indicated, whereas Figure 7.10b shows the same model with the unanalyzed relationships indicated by the curved, double-headed arrows. Note in this model that there are no curved arrows connecting endogenous variables. For example, variables D and E are expected to be correlated because they share common causes (variables A and B), but there is no curved arrow between them. The arrow is omitted because a correlation between D and E is implied by the causal structure, and it would be redundant to draw the curved two-headed arrow between them. Similarly, there is no curved two-headed arrow drawn between variables A and D because a correlation is implied by the fact that A is a cause of D. To reduce the clutter of path diagrams, such redundancies are omitted.

At this point, you should draw the curved two-headed arrows among all of your exogenous variables or, if it makes your path diagram too cluttered, omit them but put a note at the bottom of the drawing stating that all exogenous variables are assumed to be correlated.

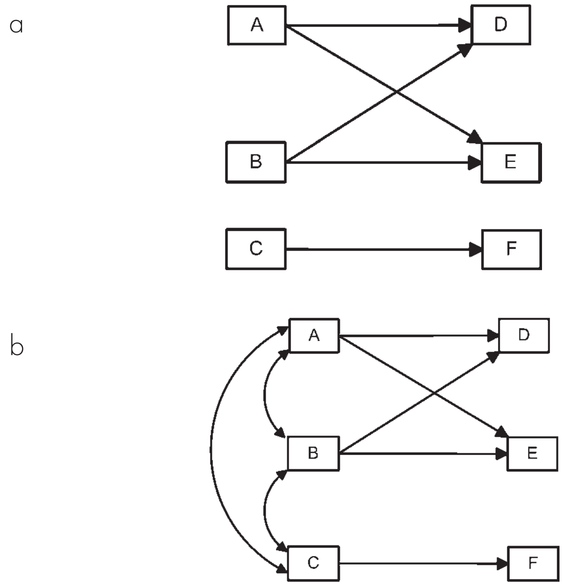


FIGURE 7.10. Examples of exogenous and endogenous variables. (a) Three exogenous and three endogenous variables; (b) unanalyzed relationships between exogenous variables.

EXPANDING THE THEORY FURTHER

Although we are almost ready to close out the theoretical system, there are some remaining details you should consider. These include temporal dynamics, disturbance terms, incorporation of a measurement theory, revisiting the existing literature, considering sign reversals, and sharing ideas with colleagues.

Temporal Dynamics

Three Types of Temporal Effects

Thus far we have assumed that the theory you have developed does not involve longitudinal features. But almost any set of variables can be examined at two points in time, three points in time, or multiple points in time. Thus, another facet you can consider adding to your theory is that of temporal dynamics. This addition can magnify the complexity of the theory considerably. To illustrate, consider a theory that consists of just one outcome and one direct cause at the same point in time. Let the cause, X, be the number of friends a child has in sixth grade and the effect, Y, be depression. The proposed theoretical relationship is that children with fewer friends are more likely to be depressed. Suppose we add a second time point, the start of seventh grade, and add the same variables to the theory at this additional point. Figure 7.11 presents a causal structure that illustrates three types of causal paths in the longitudinal model that results.

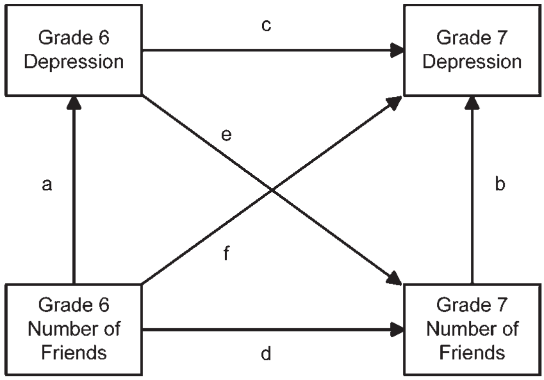


FIGURE 7.11. Models with temporal dynamics: Theory with two time points.

First, paths *a* and *b* reflect the *contemporaneous effects* of the number of friends on depression. These causal paths are the effect of *X* on *Y* within a given time period. Second, paths *c* and *d* reflect *autoregressive effects*; that is, where a variable at one point in time is assumed to influence a person's standing on that same variable at a later point in time. For example, depression in grade 6 may impact depression in grade 7, and the number of friends children have in grade 6 may influence the number of friends they have in grade 7. Finally, paths *e* and *f* reflect *lagged effects*. These are effects of a variable at time 1 on the other variable at time 2, independent of the contemporaneous and autoregressive effects that are operating. For example, the number of friends that children have in grade 6 may impact child depression in grade 7. Similarly, a child's depression in grade 6 may impact the number of friends he or she has in grade 7.

It is interesting to trace the effects of one variable on the other through the direct and indirect causal paths in this model. For example, the number of friends that the child has in grade 6 influences depression in grade 7 in four ways: (1) it impacts the number of friends the child has in grade 7, which, in turn, impacts depression in grade 7 (paths *d* and *b*); (2) it impacts depression in grade 6, which, in turn, impacts depression in grade 7 (paths *a* and *c*); (3) it impacts depression in grade 6, which, in turn, impacts the number of friends in grade 7, which, in turn, impacts depression in grade 7 (paths *a*, *e*, and *b*); and (4) it directly impacts depression at grade 7 independent of all these other effects (path *f*).

When you add a longitudinal component to your theory, you should consider modeling contemporaneous effects, autoregressive effects, and/or lagged effects between the variables. You do not need to add each of these effects. You add them only if there is a conceptual justification for doing so, and you can articulate the logic underlying them. Each of these effects can be elaborated upon using all the heuristics we have described previously.

Figure 7.12 presents a theory with three time points, grade 6, grade 7, and grade 8. This theory does not include all of the possible contemporaneous, autoregressive, and lagged effects. Nevertheless, we present it to illustrate the additional complexities with multiwave longitudinal models. For example, path *a* reflects lagged effects from a time 1 variable to a time 3 variable. Thus, one must consider not only the possible effects of variables at time $t - 1$ on variables at time t, but also the independent effects of variables at time $t - 2$ on variables at time t.

Choice of Time Intervals

In the preceding example we theorized about temporal dynamics using a 1-year interval between points. Why 1 year? Why not 6 months or 18 months? In longitudinal models, the choice of a time interval can be important. As an example, suppose a treatment to reduce child depression targets the parents of the child and teaches them more effective parenting skills for dealing with their child. The effect of the newly acquired skills on child depression will not be instantaneous. It will take time for the parents to apply them, for the child to notice a difference, and for the relationship between the parent and child to change to a positive enough state that the child starts to become less depressed. Suppose it takes a minimum of 3 months for the intervention to have its effect. Suppose further that an investigator chooses to evaluate the effects of the intervention 2 months after treatment. It will seem as if there is no treatment effect, even if the treatment has done what it was intended to do. If the researcher had waited 1 more month, an entirely different conclusion would result.

When working with longitudinal models, the choice of time intervals can influence the kinds of causal paths you include. You must carefully think through the time intervals you select to theorize about and have a rationale for the intervals upon which you ultimately settle. You should think about how long it takes for effects to manifest themselves in every longitudinal link in your theory.

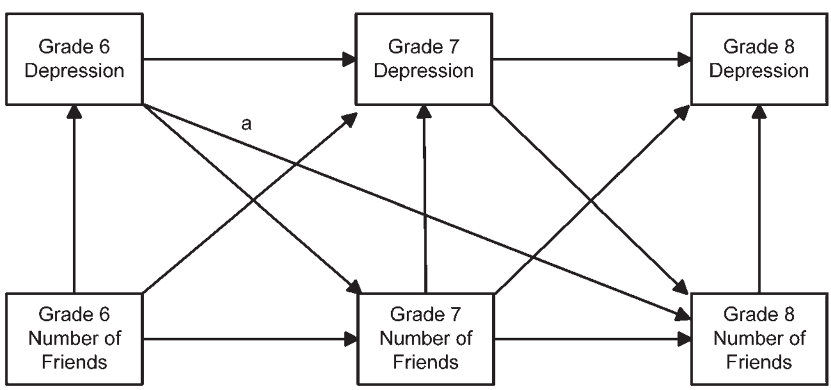


FIGURE 7.12. Three-wave theory.

Disturbance Terms

There is a subtler facet of theory construction that you can consider pursuing, although most theorists leave this to researchers who perform empirical tests of their theories. Our own preference is to be thorough and to provide researchers with a well-developed theoretical roadmap for purposes of theoretical tests, so we generally undertake this next step. But, do not expect to see it often at the level of theory description.

Consider the simple theory in Figure 7.13a. This theory has two direct causes, wherein variables *X* and *Z* are assumed to influence variable *Y*. A fourth "variable," *d*, is represented by a circle. This variable represents all unspecified variables that influence *Y* other than *X* and *Z*. This is called a *disturbance term*, and its presence explicitly recognizes that not all causal influences on a variable have been specified. Only endogenous variables have disturbance terms. Traditionally, each endogenous variable in a theory has a disturbance term associated with it.

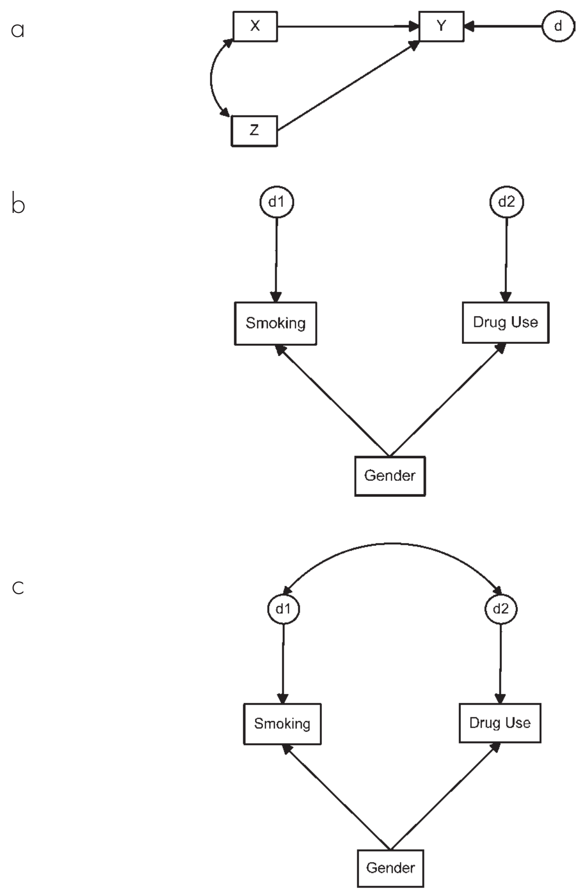


FIGURE 7.13. Examples of disturbance terms. (a) Theory with disturbance term; (b) smoking and drug example with uncorrelated disturbance terms; (c) smoking and drug example with correlated disturbance terms

Consider another example in Figure 7.13b. There are two endogenous variables and they share a common cause. One of the endogenous variables is adolescent tobacco use and the other is adolescent drug use. The common cause is gender. The theory posits that boys are more likely than girls to smoke cigarettes and that boys also are more likely than girls to use drugs. There is a disturbance term for each of the endogenous variables to acknowledge that many factors other than gender impact tobacco and drug use.

But there is a problem with this theory. According to the theory, the *only* reason that smoking cigarettes and drug use in adolescence are correlated is because they share the common cause of gender. In reality, there are many other common causes of these two constructs. For example, social class impacts both tobacco and drug use during adolescence, with more economically disadvantaged youths having an increased tendency to smoke cigarettes and to use drugs. Essentially, social class resides within the disturbance term for smoking cigarettes and for drug use. If the same unspecified cause is in each disturbance term, you would expect the two disturbance terms to be correlated. Figure 7.13c presents a more plausible theory that includes this correlation between disturbances. According to this theory, there are two reasons why adolescent cigarette smoking and adolescent drug use are correlated. One reason is because they share the common cause of gender. Another reason is that they share other common causes that are unspecified by the theory and that reside in both disturbance terms, as reflected by the presence of correlated disturbances.

A well-developed theory provides explicit statements about which disturbance terms in the framework are correlated and which disturbance terms are not. The lazy way out for a theorist is to simply assume all disturbance terms are correlated. But this is not satisfactory, and it can create considerable difficulties for testing the theory empirically. A better approach is to carefully consider every pair of disturbance terms and try to articulate a common cause that resides in each. If you can articulate such a variable, then it makes sense to posit correlated disturbances. If you cannot articulate any such variable, or if its effects are thought to be trivial, then you should not posit correlated disturbances.

For models with a longitudinal component, many theorists have a "knee-jerk" reaction that disturbances at two points in time must be correlated. Figure 7.14 illustrates a direct cause at two time points with correlated disturbances. We object to such mindless theorizing. Again, if one can articulate a compelling rationale for correlated disturbances, then by all means, correlated disturbances should be incorporated into the theory. Otherwise, correlated disturbances should be viewed with theoretical skepticism.

If you are able to articulate a variable that resides in two disturbance terms to create correlated disturbances, why not just explicitly incorporate the variable into the theoretical system? For example, for the cigarette and drug use example in Figure 7.13, why not explicitly bring social class into the theoretical system? This, of course, is the desirable route. But as in the identification of mediators, at some point we want to close out the theoretical system and work only with the variables we have specified. By including disturbance terms and correlated disturbances, we are explicitly recognizing the operation of other variables, but we choose not to give them central focus in our theory.

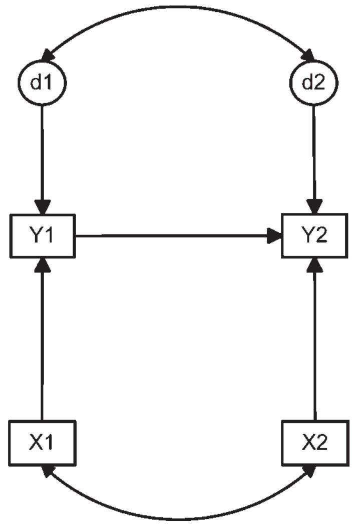


FIGURE 7.14. Example of correlated disturbances in a longitudinal model.

For your path diagram, add disturbance terms to each of your endogenous variables and then think through if correlated disturbances should be added for each pair of disturbance terms.

Latent Variables, Structural Theory, and Incorporation of a Measurement Theory

Some theorists take matters yet a step further by incorporating a measurement theory into their conceptual frameworks. This step goes beyond the province of theory construction, but we mention the general ideas here, as you may encounter them in your scientific readings. Any empirical test of a theory necessarily requires researchers to develop and use measures of the theoretical constructs. Just as one can build a theory linking one concept to another concept, so too can one build a theory linking a construct to a measure of that construct. Some theorists combine both types of theories into a single overarching framework.

A measurement theory makes a distinction between a latent variable and an observed measure of that variable. The latent variable is the true construct that you are interested in making statements about—for example, depression. Although we can see the symptoms and overt manifestations of depression, we can't directly observe the seat of depression in a person's mind. Instead, we rely on some observable response(s) to assess the latent variable, such as a multi-item inventory of depression that a person might complete. Figure 7.15a contains one representation of a measurement model. The latent variable of depression is contained in a circle, and the observed measure thought to reflect depression is contained in a square (the label *AR* stands for *adolescent report* of depression). A causal path is drawn from the latent variable to the observed measure, under the assumption that how depressed a person is influences how he or she responds to the questions on the inventory. There also is an error term, (*e*), that reflects measurement error; that is, there are factors other than depression that influence a person's responses on the inventory. Ideally, measurement error is minimal, but it is a fact of life for many research endeavors. The relationship between the latent construct and the observed indicator is usually assumed to be linear, but it could also be nonlinear.

Sometimes we obtain multiple indicators of a construct. For example, a researcher might obtain a self-report of depression from an adolescent as well as a report from the adolescent's mother about how depressed the child is (*MR*). A measurement model for this scenario is presented in Figure 7.15b. The latent variable of depression is assumed to influence both of the observed measures, and each measure is assumed to have some measurement error as reflected by the presence of error terms. The errors are assumed to be uncorrelated because we cannot articulate any viable reason why we would expect

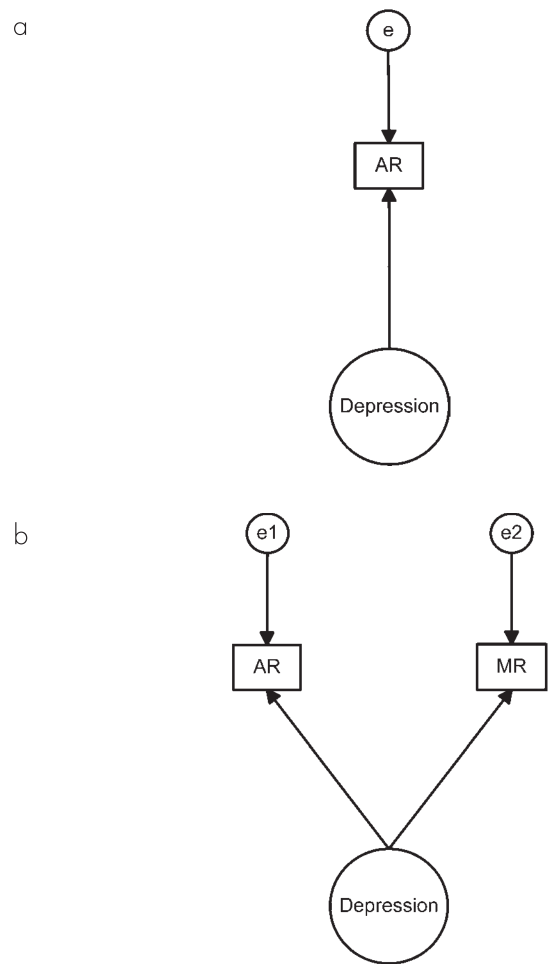


FIGURE 7.15. Measurement models. (a) Single indicator; (b) multiple indicators.

them to be correlated. However, one can introduce correlated measurement error, if appropriate.

Figure 7.16 presents an example of a more elaborate theoretical framework that incorporates a theory about the relationship between constructs as well as a measurement theory. Although it appears somewhat intimidating, it is a straightforward model. There are five latent constructs, and the main substantive theory is focused on them. The portion of the diagram focused on the causal relations among the latent variables is called the *structural model*. The primary outcome variable in this model is the birthweight of a newborn. Birthweight is thought to be influenced by two factors: how much alcohol the mother consumes during her pregnancy and how much she smokes during her pregnancy. Both of these variables are thought to be influenced by two other variables. The first determinant is the extent of support the mother receives from friends and relatives who can help her quit smoking and drinking. The second is the mother's locus of control. *Locus of control* refers to the extent to which the mother believes that what happens to her is beyond her control. The theory is that the more a mother thinks that what happens is not under her control, then more likely she will be to keep smoking and drinking during pregnancy. These two latent exogenous variables are assumed to be correlated. Each of the three latent endogenous variables has a dis-

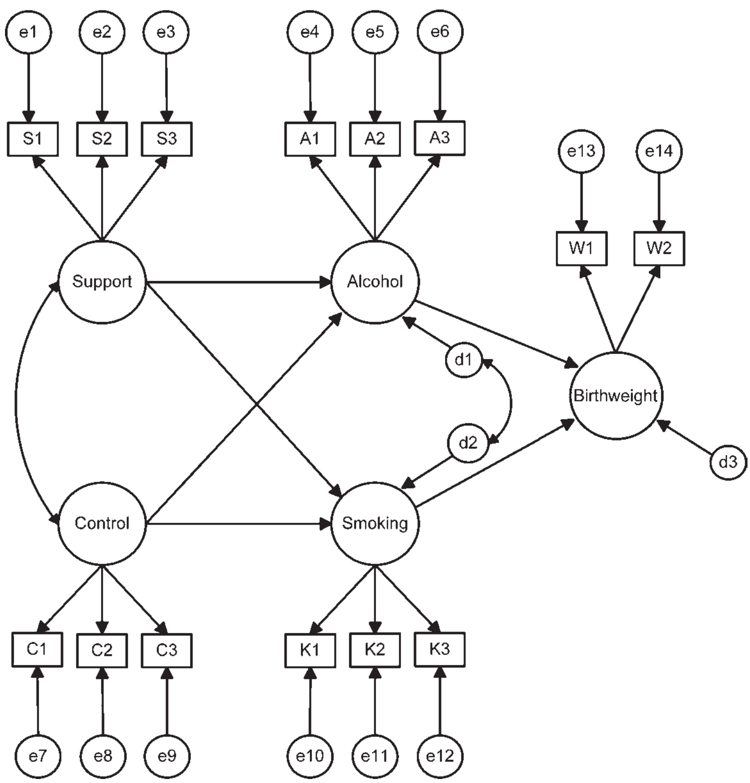


FIGURE 7.16. Example of integrated structural and measurement model.

turbance term, indicated by a circle with a *d* inside of it. The disturbances are assumed to be correlated for alcohol use and smoking.

The portion of the diagram with arrows from the latent constructs to the observed measures constitutes the *measurement model*. Each of the latent variables has multiple indicators; that is, the researcher obtains three measures of each construct, except birthweight, which is measured using two different indicators. In the interest of space, we do not describe these measures, but note that each is assumed to be fallible, that is, subject to some measurement error (see the circles ranging from e1 to e14). The measurement errors are assumed to be uncorrelated. All of the relationships in the model are assumed by the theorist to be linear in form. Figure 7.16 provides an explicit roadmap for a researcher to test the combined structural theory and measurement theory.

We will not ask you to incorporate a theory of measurement into your path diagram, as this goes beyond the scope of this book.

Revisiting Your Literature Review

Before closing out the theoretical system, you want to revisit all of the relevant scientific literature on your outcome variable and the other variables included in your system that you read before embarking on the theory construction enterprise. In relation to this literature, which variables have you included in your theory that the literature has failed to include? These represent innovations on your part. What relationships have you elucidated that the literature has failed to elucidate? These also represent potential for a new contribution to scientific knowledge. What variables has the literature suggested that are omitted from your theory? You should consider bringing these into your theory. What relationships has the literature established that you have not or that contradict what you have theorized? Note these and make adjustments to your theory accordingly.

Some Final Steps

There are two additional steps we recommend you pursue. First, take every direct relationship you have specified and try reversing its sign. That is, if the relationship is positive, try making it inverse and see if you can articulate a logic that would justify this reversal. If the relationship is inverse, try making it positive and see if you can articulate a logic that would justify this reversal. If you are able to articulate compelling logic for both a direct and an inverse relationship, then you essentially have articulated competing theories that lead to opposite predictions. It is then an empirical question as to which theory is correct. As you consider relationship sign reversals, new mediators or moderators might come to mind, and you may want to alter your theory accordingly.

As a final step, show your theory to friends and colleagues and discuss it with them. Ask them what they think about it. Do they agree or disagree with it? Can they suggest variables you have left out or variables you should drop? Pursue input from diverse sources. When this is done, close out the theoretical system.

BOX 7.1. Finding Sources for a Literature Review

A major strategy for developing ideas about causal models is reading about research that has been conducted in the area in which you are working. There are several methods for locating relevant literature. One procedure involves the use of computer searches of scientific journals and books, which are available in most college libraries (e.g., PsycINFO, Medline). In this procedure, you specify a set of "keywords" to search. For example, if you are studying attitudes toward abortion, you might do separate searches on the keywords *abortion*, *attitudes toward abortion*, or *pregnancy resolution*. The computer then scans the titles and abstracts of a large number of scientific journals, and a list of the titles and abstracts of all relevant articles that contain the keywords is provided. Check with your librarian for details about accessing these databases and conducting an electronic search.

In our experience, a computer search is only as good as one's ability to generate a good list of keywords. The results of such a search may miss important articles because the author of an article did not use one of your keywords in the abstract or title. Also the search can include a good number of irrelevant articles. We often search first on an obvious keyword and then scan the abstracts of the "hits" to get ideas for additional keywords. We then follow up the initial search with more searches based on these new keywords.

A second approach to identifying relevant literature is called the "grandfather method." In this approach, you first identify scientific journals where relevant articles are likely to have been published (this list can frequently be generated with the help of a professor or some other "expert," as well as the above computer search strategy). You then go to the Table of Contents of each issue of each journal for the past 5 years and identify articles that seem relevant based on their titles and abstracts. If an article is deemed relevant, you secure a copy of it, read it, and then examine its reference section for additional relevant articles, based on what you read. Then you locate these cited articles and repeat the process for each of them. The result will be a set of articles that appeared in the major journals and articles that were cited by these articles. The key to this method is to examine the bibliography of every article you locate, to further identify relevant research.

Another approach for identifying relevant literature is to use the Science Citation Index and/or the Social Science Citation Index. These are reference books or databases contained in most college libraries; they list, for a given author of a given paper, all of the articles published by other individuals who have cited the paper (an online version of both indices also exists). If you are aware of the author of a major article in the area in which you are working, then the citation index can be a useful way of identifying other researchers who have cited that article in the context of their published research. The relevant publications of

cont.

these other researchers can then be identified by information provided in the citation index.

Another strategy that can augment the above is to use the Internet to locate the websites of scientists who have published in the area in which you are working. Many professors and applied scientists maintain websites, on which they post their most recent research papers for downloading, some of which have not yet been published.

A final strategy is to use Google Scholar, a specialized search tool developed by the website Google for identifying papers and articles that have cited other papers and articles (see the Google website for details at *www.google.com*).

Once you have identified the relevant literature, read it! Don't simply look at summaries of the research.

PERSPECTIVES ON THE CONSTRUCTION OF CAUSAL THEORIES

Path Diagrams as Theoretical Propositions

What started as a fairly simple theory and path diagram have probably now blossomed into a fairly elaborate theoretical network. An invaluable aid to developing the theory has been the path diagram that we continually updated, elaborated, and expanded. Many theories in the social sciences are simple three- or four-variable systems consisting only of direct causal relationships. Such theories are straightforward to describe using narratives, and it is easy to keep in mind the overall framework the theorist is elaborating. However, as theories grow in complexity, readers may need some type of pedagogical device to help them see the broader framework in a unified way. Path diagrams are useful in this regard. A path diagram summarizes many theoretical propositions that, if expressed verbally, would constitute a long list. Every causal path in a path diagram represents a theoretical proposition, and the absence of causal paths also can reflect theoretical propositions, such as propositions about complete mediation as opposed to partial mediation. To illustrate, consider the structural model in Figure 7.16. Here are the major theoretical propositions that derive from this path diagram:

Proposition 1: The birthweight of a newborn is influenced by how much a mother smokes during pregnancy. The more a mother smokes during the pregnancy, the lower the birthweight of the newborn. This relationship is assumed to be linear.

Proposition 2: The birthweight of a newborn is influenced by how much alcohol a mother consumes during pregnancy. The more alcohol a mother consumes during the pregnancy, the lower the birthweight of the newborn. This relationship is assumed to be linear.

Proposition 3: The amount a mother smokes during her pregnancy is influenced by the extent of her support network for quitting smoking. The more support the mother has to quit smoking, the less she will smoke during her pregnancy. This relationship is assumed to be linear.

Proposition 4: The amount a mother smokes during her pregnancy is influenced by her locus of control. The higher the locus of control, the less she will smoke during her pregnancy. This relationship is assumed to be linear.

Proposition 5: The amount of alcohol a mother consumes during her pregnancy is influenced by the extent of her support network for quitting drinking. The more support the mother has to quit drinking, the less she will drink during her pregnancy. This relationship is assumed to be linear.

Proposition 6: The amount of alcohol a mother consumes during her pregnancy is influenced by her locus of control. The higher the locus of control, the less she will drink during her pregnancy. This relationship is assumed to be linear.

Proposition 7: The effects of locus of control on birthweight are completely mediated by how much a mother drinks and how much a mother smokes.

Proposition 8: The effects of the support network on birthweight are completely mediated by how much a mother drinks and how much a mother smokes.

Proposition 9: The correlation between how much a mother drinks and how much she smokes during pregnancy is a function of the common causes of locus of control and the extent of support network.

These propositions omit statements about correlated errors and the measurement model.

Years ago, when submitting grant proposals to secure funding to conduct research, it was common practice to list the specific aims and formal hypotheses early in the proposal and then to coordinate discussion of the literature, elaboration of measures, and specification of data collection and data analysis around the three or four theoretical propositions stated in the aims section. This also was a common practice in scientific reports, where the introduction of the report would culminate in the formal statement of three or four hypotheses. However, as theory becomes multivariate and complex, which is more often the case in modern-day social science, these traditions become inefficient and detract from effective communication. Path diagrams can be a useful tool for summarizing theoretical propositions efficiently. Each path in the diagram can be labeled with a + or a – to indicate if the presumed relationship is assumed to be positive or inverse. Nonlinear relationships can be described in the text, either verbally or mathematically (using principles discussed in Chapter 9).

Unfortunately, some scientists fail to appreciate that path diagrams are the essence of multiple hypotheses and theoretical propositions. Thus, you may receive criticism in a grant proposal or a research report for not formally stating specific hypotheses, despite the fact that you have presented a clear and explicit path diagram.

Another potential problem with the use of path diagrams comes from the opposite end of the spectrum. Some reviewers of proposals and reports do not believe you have a theory unless you have presented a path diagram. We have served on numerous review panels and have observed instances where a research project is said to "lack theory," only to see a similar project move forward uncriticized simply because it had a diagram with boxes and arrows. The variables in the path diagram were abstract and fuzzy, the posited relationships were not well thought out or articulated, and crucial variables were omitted, but because there were boxes and arrows, the research was deemed as having a viable theoretical base.

We raise these issues so that you will not get discouraged if you are criticized for not specifying hypotheses after having presented a path diagram and so you will not be lackadaisical and think you can get by with any path diagram. If you use the heuristics we have described in this chapter and previous ones, if you carefully focus your concepts and relationships, and if you can articulate the logic underlying every path in your path diagram, you should be on sound theoretical footing.

A Note on Research Design and Statistical Analysis

Theories of the form we have developed in this chapter often are associated with a method of statistical analysis called *structural equation modeling* (SEM). SEM tends to be used with research designs that are correlational or observational in nature rather than experimental, though SEM is also readily applied to experimental data. The theory construction strategies described in this chapter are useful independent of the type of research one uses to test theories or how one chooses to analyze data. If you think in terms of cause and effect, then the ideas presented in this chapter are relevant.

As an example, many areas of the social sciences develop and evaluate interventions designed to impact healthy or unhealthy behavior. Research on interventions typically includes a treatment group that receives the intervention and a control group that does not. In such cases, the outcome variable (e.g., social anxiety, exercise, drug use) is the "effect," and the "cause" is the presence or absence of the intervention (where the treatment group represents "the presence of the intervention" and the control group represents "the absence of the intervention"). In trying to understand the effects of the intervention, the social scientist builds a theory of the relationship between these two variables. Usually, the social scientist begins by assuming a direct causal relationship between them. From this juncture, he or she then starts the theory construction process of specifying mediators and moderators, as described throughout this chapter.

Elaborating the Logic Underlying Each Path

Just as grounded and emergent theorists need to build an effective case for each conclusion or thesis they present, so too must the theorist who posits causal mechanisms using the approach described in this chapter do so. Each path in a path diagram is a "thesis" or a "potential conclusion," and the theorist needs to build as strong as case as possible for the underlying a priori logic. In Chapter 10 we describe theories of argumentation and rhetoric to provide perspectives that grounded and emergent theorists consider in describing the bases for their theses and conclusions. These same techniques can be used when building a case for a given path in a causal theory.

The Use of Causal Analysis in Grounded/Emergent Theorizing

The development of theory using the methods described in this chapter has emphasized an a priori approach to theory construction. However, there is no reason why the concepts that we have developed cannot be applied to grounded and emergent theory construction as well. Specifically, the grounded/emergent theorist can approach the analysis and interpretation of qualitative data by constructing a path diagram that captures conceptually the causal relations among variables that have been judged to emerge from the data. In framing and thinking about the data, the theorist can think about direct causes, indirect causes, partial and complete mediation, moderated relationships, bidirectional relationships, and spurious relationships; he or she can think about mediated mediation, moderated mediation, mediated moderation, and moderated moderation. In short, the causal framework can be used as a blueprint for the types of relationships that grounded/emergent theorists think about as they approach the theory construction process from qualitative data. Indeed, most computer software for the analysis of qualitative data and for the development of grounded theory contains features for drawing path diagrams that capture the presumed causal dynamics in the qualitative data, though not in the detail represented in this chapter.

SUMMARY AND CONCLUDING COMMENTS

We conclude by listing the sequence of steps you might pursue in constructing your theory if you adopt a causal framework:

- Step 1: Identify the outcome variable in which you are interested.
- Step 2: Using heuristics from throughout this book, identify two or three direct causes of the outcome.

  (*Note*. The above two steps could instead be (1) specify a variable in which you are interested and (2) identify two or three consequences or outcomes associated with that variable. Or specify a single direct cause consisting of an intervention and an outcome. Whatever the sequence, you want to have at least one direct effect in your theory at this juncture.)
- Step 3: Turn the direct causes into mediated effects using either the *why* heuristic or the *cause of a cause* heuristic.
  - Step 3a: For each mediator, specify complete or partial mediation.

- Step 4: For every direct effect in the model, consider adding a moderator variable. Focus each moderated relationship using a hypothetical mean factorial table (per Chapter 6).
- Step 5: Expand and refine the mediated and moderated portions of the model.
  - Step 5a: For every mediated effect in the model, consider adding moderated mediation.
  - Step 5b: For every moderated effect in the model, consider adding mediated moderation.
  - Step 5c: For every moderated effect in the model, consider adding moderated moderation.
- Step 6: For every direct effect in the model, consider adding reciprocal causation.
  - Step 6a: For every reciprocal causal effect in the model, consider adding moderated reciprocal causation.
  - Step 6b: For every moderated reciprocal causation effect, think about adding moderated moderation.
  - Step 6c: Consider turning a reciprocal causal effect into a feedback loop by adding mediators.
  - Step 6d: If feedback loops are added with mediators, consider the issue of partial versus complete mediation.
- Step 7: Consider adding new outcome variables to the system.
- Step 8: Consider adding effects of effects in order to turn outcomes into mediators. Specify partial or complete mediation for new "mediators" and consider adding mediated moderation, moderated mediation, and moderated moderation to this part of the system.
- Step 9: Consider adding temporal dynamics to the model, including contemporaneous effects, autoregressive effects, and lagged effects.
- Step 10: Fine-tune the relationships and logic of your model.
  - Step 10a: Allow all exogenous variables to be correlated.
  - Step 10b: Add disturbance terms for all endogenous variables and consider the need to add correlated disturbances.
  - Step 10c: Focus all concepts and all relationships using strategies from Chapters 5 and 6.
  - Step 10d: Revisit your initial review of the literature and make changes to the theory, as appropriate. Flag innovations in your theory.
  - Step 10e: Consider sign reversals for all direct relationships in your theory.
  - Step 10f: Get feedback from colleagues.

The 10-step sequence may make little sense for the grounded theorist. Rather, the different components of the causal model are pieced together through systematic analysis of the qualitative data and field notes. We return to this issue for grounded theorists in Chapter 10.

SUGGESTED READINGS

Blalock, H. M. (1964). *Causal inferences in non-experimental research*. Chapel Hill: University of North Carolina Press.—An excellent discussion of the concept of causality and methods of theory construction.

Blalock, H. M. (1983). *Theory construction*. San Francisco: Jossey Bass.—A discussion of general methods of theory construction.

Bunge, M. (1961). Causality, chance, and law. *American Scientist,* 69, 432–488.—A clear discussion of aspects of causality.

Cartwright, N. (2007). *Hunting causes and using them*. New York: Cambridge University Press.—A more advanced discussion of the concept of causality, particularly as applied in the field of economics.

Howard, G. S. (1984). The role of values in the science of psychology. *American Psychologist,* 40, 255–265.—A discussion of how values influence the way we theorize.

Manicas, P. (2006). *A realist philosophy of social science: Explanation and understanding*. Cambridge, UK: Cambridge University Press.—A more advanced discussion of causal realism and a critique of traditional views of causality.

Maxwell, J. (2004). Using qualitative methods for causal explanation. *Field Methods,* 16, 243–264.—An introduction to alternative views of causality as used in grounded theories.

Morgan, S., & Winship, C. (2007). *Counterfactuals and causal inference*. New York: Cambridge University Press.—A lucid discussion of the use of counterfactuals in the analysis of causality.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. New York: Cambridge University Press.—A comprehensive and lucid discussion of causality from many perspectives, philosophical and otherwise.

Salmon, W. C. (1998). *Causality and explanation*. New York: Oxford University Press.—A discussion of the approach of causal realism.

Zetterbert, H. (1965). *On theory and verification in sociology*. Totowa, NJ: Bedminster Press.—A discussion of strategies for theory construction using the concept of causality.

KEY TERMS

predictive relationship (p. 138)	criterion variable (p. 138)
predictor variable (p. 138)	causal relationship (p. 139)

outcome variable (p. 141)
path diagram (p. 141)
independent variable (p. 141)
dependent variable (p. 141)
determinant (p. 141)
direct causal relationship (p. 141)
indirect causal relationship (p. 142)
mediated relationship (p. 142)
spurious relationship (p. 142)
mediator (p. 142)
moderated causal relationship (p. 143)
bidirectional causal relationship (p. 143)
reciprocal causal relationship (p. 143)
unanalyzed relationship (p. 143)
moderator variable (p. 143)
endogenous variable (p. 145)
exogenous variable (p. 145)
partial mediation (p. 147)
complete mediation (p. 147)
why heuristic (p. 147)
partial mediator (p. 148)
cause of a cause heuristic (p. 149)
mediated moderation (p. 151)
moderated mediation (p. 151)
moderated moderation (p. 152)
stronger-than heuristic (p. 153)
equilibrium assumption (p. 153)
feedback loops (p. 154)
moderated reciprocal causation (p. 155)
contemporaneous effects (p. 160)
autoregressive effects (p. 160)
lagged effects (p. 160)
disturbance term (p. 162)
correlated disturbances (p. 163)
latent variable (p. 164)
structural model (p. 166)
measurement model (p. 167)

EXERCISES

Exercises to Reinforce Concepts

1. Distinguish between causal and predictive relationships.
2. What are the five common features of the construct of causality upon which most social scientists agree?
3. Identify and define the six basic types of relationships in causal models and give an example of each.
4. What is the essence of a causal relationship? Why have some philosophers objected to the notion of causality?
5. If causality can never be proven, why is the concept still useful in science?
6. What is a path diagram?
7. What strategies or heuristics can you use to turn a direct relationship into an indirect relationship? Create an example using them.

8. What is the difference between partial and complete mediation?
9. What heuristics do you use to identify moderated relationships?
10. What is the difference between mediated moderation, moderated mediation, and moderated moderation?
11. How are feedback loops indirect effects?
12. Why is there no such thing as an instantaneous reciprocal causal relationship?
13. What heuristics might lead to the addition of spurious effects in a theory?
14. Are spurious effects always bad? Why or why not?
15. What is the difference between an exogenous and an endogenous variable?
16. What are the three types of relationships that incorporate temporal dynamics into them?
17. How is the time frame important for analyzing mediation?
18. Under what conditions do you specify correlated error?
19. What is the difference between a structural model and a measurement model?

Exercises to Apply Concepts

1. Find a study in the literature and describe the theory it tests using a causal framework. Draw a path diagram of the theory. Provide conceptual definitions of each construct and be explicit about the nature of each relationship in the theory.
2. Using the principles discussed in this chapter, construct a causal theory. Include a path diagram of it and an accompanying narrative describing it. Give precise and clear conceptual definitions of each variable, using the strategies in Chapter 5. Clarify the relationships, as necessary, using the thought experiment strategies in Chapter 6.

8

Mathematical Modeling

Even if there is only one possible unified theory, it is just a set of rules and equations.

—STEPHEN W. HAWKING (1988)

This chapter describes an approach to theory construction called *mathematical modeling*. Like causal modeling, the approach involves describing relationships between variables, but the emphasis is on describing those relationships using mathematical concepts. Mathematical models can be used in conjunction with causal thinking, as we demonstrate in a later section of this chapter, but social scientists who use mathematical modeling tend not to think in terms of indirect causes, mediated relationships, and moderated relationships in the way that we outlined in Chapter 7. Instead, they focus on thinking about functions and describing relationships mathematically based on functions. They more often than not use nonlinear relationships. Our focus here is not on integrating causal and mathematical modeling as approaches to theory construction. Rather, we merely wish to provide you with an additional tool for your theory construction toolbox, mathematical modeling, as you strive to gain insights into the phenomena you want to study.

Constructing mathematical models can involve complex mathematics that go well beyond the background of many readers of this book. Entire books have been written on mathematical modeling that assume years of study of calculus and formal mathematics. Our treatment must, accordingly, be limited, and we provide only a general sense of building mathematical models and thinking as a math modeler would. However, the chapter should be a good starting point for delving into this approach in more depth vis-à-vis the suggested readings at the end of the chapter.

Mathematical modeling is common in the physical sciences, but it is used less often in the social sciences. Our goal is to provide you with a sense of mathematical modeling as it is pursued in the social sciences. In this chapter we first expose you to basic concepts and terms you will encounter as you read about math models or pursue mathematical modeling. More specifically, we distinguish between categorical, discrete, and continuous variables; differentiate axioms and theorems; introduce the notion of a function; use linear functions to identify key features of functions; and describe the difference between deterministic and stochastic models. We also provide an intuitive overview of derivatives, differentiation, integrals, and integration in calculus, as well key notions of model identification and metrics. We next describe five commonly used functions in math models: logarithmic functions, exponential functions, power functions, polynomial functions, and trigonomic functions, as well as functions for categorical variables. We conclude our background section by considering ways of transforming and combining functions and building functions for multiple variable scenarios.

Following the presentation of these key concepts, we describe the phases of building a mathematical model and then provide four examples of such models in the social sciences. We then briefly characterize chaos theory and catastrophe theory as influential mathematical models in the social sciences. Our initial discussion may seem a bit fractured as we develop one mathematical concept after another. Be patient. Later sections will pull it all together.

TYPES OF VARIABLES: CATEGORICAL, DISCRETE, AND CONTINUOUS

In Chapter 6 we distinguished between categorical and quantitative variables. A categorical variable has different “levels,” “values,” or “categories,” and there is no special ordering to the categories along an underlying dimension. The categories are merely labels that differentiate one group from another (e.g., “male” or “female” for the variable of gender). In contrast, a quantitative variable is one in which individuals (in the social sciences) are assigned numerical values to place them into different categories, and the numerical values have meaning in that they imply more or less of an underlying dimension that is of theoretical interest.

Mathematical modelers make distinctions between discrete quantitative variables and continuous quantitative variables. A *discrete variable* is one in which there are a finite number of values between two values. For example, for the number of children in a family, there is a finite number of values, say, between 1 child and 4 children, namely the values of 2 children and 3 children. We do not think of there being 1.5 children or 1.7 children in a family. For a *continuous variable*, however, there is an infinite number of values between any two values. Reaction time to a stimulus is an example of a continuous variable. Even between the values of 1 and 2 seconds, an infinite number of values could occur (e.g., 1.001 seconds, 1.873 seconds, 1.874 seconds).

Whether a variable is classified as discrete or continuous depends on the nature of the underlying theoretical dimension and not on the scale used to measure that dimension. Tests that measure intelligence, for example, yield scores that are whole numbers (e.g., 101, 102); hence the scores are discrete. Nevertheless, intelligence is continuous in nature because it involves a dimension that permits an infinite number of values to occur, even though existing measuring devices are not sensitive enough to make such fine distinctions. In the reaction time example, the measurement of time can be very precise with modern equipment, but there even is a limit to the precision possible with measures of time. Such limits in the precision of measurement do not make the underlying dimension discrete. Reaction time is continuous in character.

Even though social scientists often must rely on discrete measures of continuous constructs, they build models with those measures as if they were continuous. As long as the measures are comprised of many values, this practice usually is not problematic. It is only when the number of values in the measure of a continuous variable is few that problems can arise and special considerations in the modeling effort need to be made.

The distinction between discrete and continuous variables is important because the strategies used to construct a mathematical model differ depending on whether the quantitative variables are discrete or continuous. We devote most of our attention to the case in which the theorist is working with continuous variables, but we occasionally consider qualitative and discrete variables as well.

AXIOMS AND THEOREMS

The term *axiom* is used in many ways in the social sciences, but in mathematics, an axiom is a mathematical statement that serves as a starting point from which other mathematical statements are logically derived. Axioms are "given." They are not derived through deduction, nor are they the subject of mathematical proofs. They are starting points. By contrast, a *theorem* is a statement that can be logically derived from, or is proven by, one or more axioms or previous statements. The use of these terms and many variants of them (e.g., a proposition, a lemma, a corollary, a claim, an identity, a rule, a law, a postulate, a principle) vary somewhat depending on the branch of mathematics, but the above characterization captures the essence of axioms and theorems as used in mathematical models in the social sciences.

FUNCTIONS

Functions are central to mathematical modeling. A simple analogy for thinking about functions is to think of a machine that you put something into and get something back, based on what you input. For example, you might press a key that inputs the number 3 into a machine and out comes the number 9. You might press another key that inputs the number 5 into the machine and out comes the number 15. The machine in this case represents the function "take the input value and triple it." Functions in math typically involve numbers as inputs and outputs.

Suppose we decide to name our machine *Jack*. We can write the function that the machine performs as follows:

$$\text{Jack}(X) = 3X$$

This equation states that whatever the value of X, the value of "Jack of X" will be triple it. The traditional notation is to name the machine f (for "function") and write it as follows:

$$f(X) = 3X$$

All functions have what are called a domain and a range. The *domain* is the set of possible input values, and the *range* is the set of possible output values. The domain and range often are stated mathematically rather than listed as individual numbers. For example, for the function

$$f(X) = \sqrt{X - 3}$$

the domain or possible input values is any number greater than or equal to 3 (because you can not calculate the square root of a negative number), and the range or possible output values is any value greater than or equal to 0.[1] The domain is any number that produces a "meaningful output" and that will not cause the machine to malfunction (e.g., the number 2, which would require us to calculate the square root of –1). A shorthand way that mathematical modelers use to express the domain is "the domain is $\{X|X \geq 3\}$," where the symbol "|" is read as "given that." This expression states that the domain is equal to X, given that X is greater than or equal to 3. This may seem a bit cryptic, but it is an efficient way of stating a domain or a range. For example, I might have a function where the domain is $\{X|X > 0\}$ and the range is $\{Y|Y > 0\}$.

Functions can apply to more than a single input. For example, the function $f(X,Z) = X - Z$ has two inputs, X and Z, and an output that is the difference between them. For example, if $X = 5$ and $Z = 2$, the function $f(X,Z)$ yields the output 3.

Functions are the foundation of mathematical models. When one "maps" a function between Y and X, one attempts to specify what function applied to values on X will produce the values of Y. A central task in mathematical modeling is that of mapping functions.

LINEAR FUNCTIONS

One of the most commonly used functions in the social sciences is the linear function. In this section we describe the nature of linear functions and then use them to illustrate

[1]For pedagogical reasons, we restrict all examples in this chapter to real numbers.

basic issues in building mathematical models. In later sections we consider other functions.

The Slope and Intercept

The Slope

Consider the two-variable example we used in Chapter 6 to develop the general idea of a linear relationship, namely, the number of hours an employee worked, *X*, and the amount of money paid to the employee, *Y*. Assume a scenario where each of four individuals works at a rate of $1 per hour. Their scores are:

Individual	*X* (hours worked)	*Y* (dollars paid)
1	1	1
2	4	4
3	3	3
4	2	2

The relationship between *X* and *Y* is illustrated in Figure 8.1, which uses a scatterplot with connected dots. As indicated by the straight line on the scatterplot, there is a linear relationship between *X* and *Y*. This relationship can be stated mathematically as

$$Y = X$$

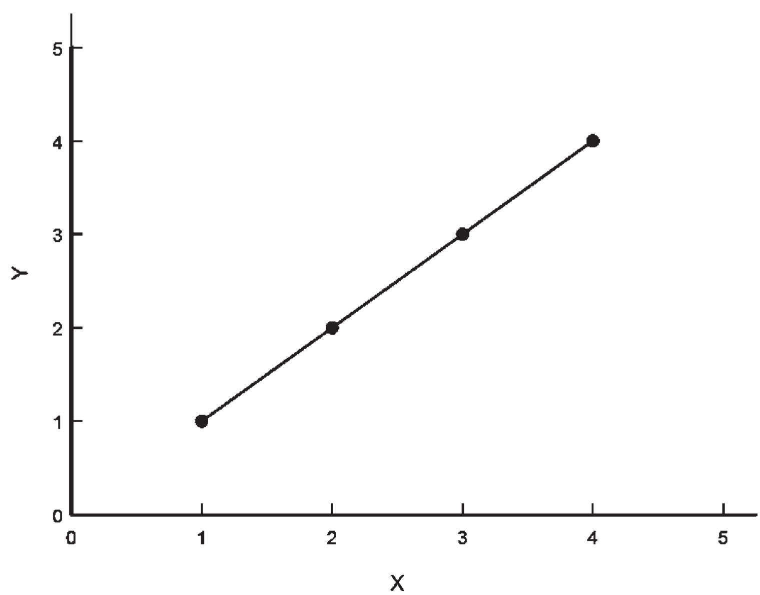


FIGURE 8.1. Linear relationship with slope = 1.

In other words, the number of dollars paid equals the number of hours worked.

Suppose the individuals were not paid $1 per hour, but instead were paid $2 per hour. The scores on *X* and *Y* would be as follows:

Individual	*X* (hours worked)	*Y* (dollars paid)
1	1	2
2	4	8
3	3	6
4	2	4

In this case, the relationship between *X* and *Y* can be stated as

$$Y = 2.00X$$

In other words, the number of dollars paid equals 2 times the number of hours worked. Figure 8.2 presents a scatterplot of these data (line *B*) as well as the data from Figure 8.1 (line *A*). (Line *C* is explained on p. 184.) Notice that we still have a straight line (and, hence, a linear relationship) but, in the case of $2 per hour, the line rises faster than with $1 per hour; that is, the slope of the line is now steeper. Technically, the slope of a line indicates the number of units that variable *Y* changes when variable *X* increases by 1 unit. It is the rate of change in *Y* given a 1-unit increase in *X*. When the wage is $2 per hour, a person who works 1 hour is paid $2, a person who works 2 hours is paid $4,

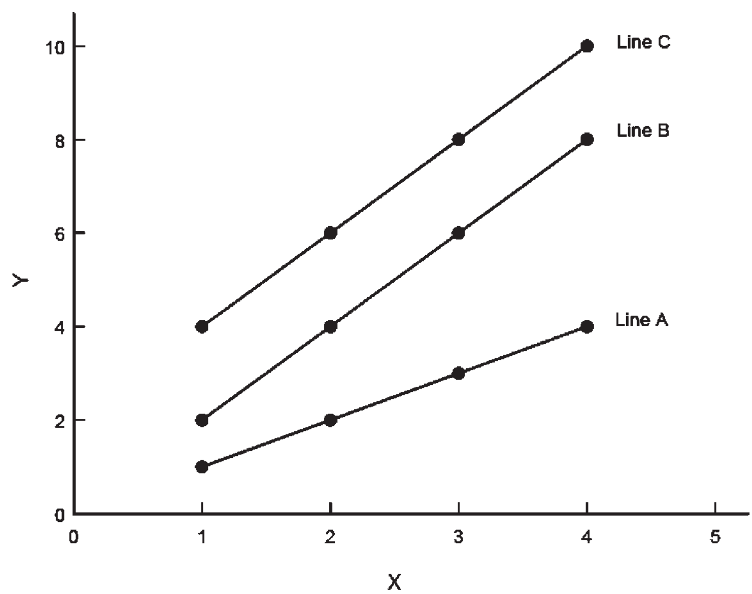


FIGURE 8.2. Example of linear relationships with different slopes or intercepts.

and so on. When X increases by 1 unit (e.g., from 1 to 2 hours), Y increases by 2 units (e.g., from \$2 to \$4). The slope that describes this linear relationship is therefore 2. In contrast, the slope that describes the linear relationship $Y = X$ is 1.0, meaning that as X increases by 1 unit, so does Y. One way in which linear relationships differ is in terms of the slopes that describe them.

The slope that describes a linear relationship can be determined from a simple algebraic formula. This formula involves first selecting the X and Y values of any two individuals. The slope is computed by dividing the difference between the two Y scores by the difference between the two X scores; in other words, the change in Y scores is divided by the change in X scores. Algebraically,

$$b = (Y_2 - Y_1)/(X_2 - X_1) \tag{8.1}$$

where b represents the slope, X_1 and Y_1 are the X and Y scores for any one individual, and X_2 and Y_2 are the X and Y scores for any other individual. In our example, inserting the scores for individuals 1 ($X = 1$, $Y = 2$) and 2 ($X = 4$, $Y = 8$) into Equation 8.1, we find that the slope for line B is

$$b = (2 - 8)/(1 - 4) = 2.00$$

This is consistent with what was stated previously.

The value of a slope can be positive, negative, or 0. Consider the following scores:

Individual	X	Y
1	2	3
2	1	4
3	4	1
4	3	2

Inserting the scores for individuals 2 and 4 into Equation 8.1, we find that the slope is

$$b = (4 - 2)/(1 - 3) = -1.00$$

Figure 8.3 presents a scatterplot of the data for this relationship. The relationship is still linear, but now the line moves downward as we move from left to right on the X axis. This downward direction characterizes a negative slope, whereas an upward direction characterizes a positive slope. A slope of 0 is represented by a horizontal line because the value of Y is constant for values of X.

In sum, a positive slope indicates a *positive* or *direct linear relationship* between X and Y, whereas a negative slope indicates a *negative* or *inverse linear relationship* between X and Y. In the case of a positive relationship, as scores on X *increase*, scores on Y also *increase*. In the case of a negative relationship, as scores on X *increase*, scores on Y

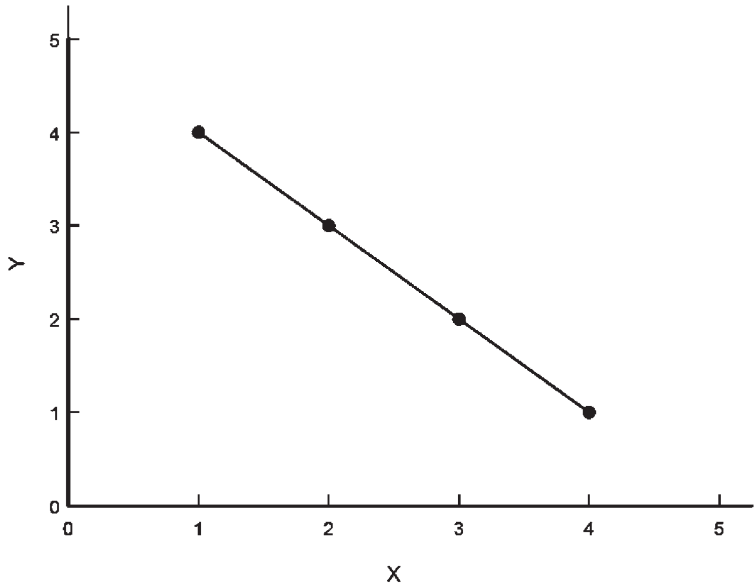


FIGURE 8.3. Example of a negative slope.

decrease. For instance, the slope in Figure 8.3 is –1.00, meaning that for every unit *X* increases, *Y* decreases by one unit.

The Intercept

Let us return to the example where individuals are paid $2 per hour worked. Suppose that in addition to this wage, each individual is given a tip of $1.50. Now the relationship between *X* and *Y* is

$$Y = 1.50 + 2.00X \tag{8.2}$$

Line *C* of Figure 8.2 plots this relationship for the four individuals. If we compute the slope of this line, we find it is 2.00, as before. Notice that lines *C* and *B* are parallel but that line *C* is higher up on the *Y* axis than line *B*. The amount of separation between these two lines can be measured at the *Y* axis, where $X = 0$. When $X = 0$, the *Y* value is 1.50 for line *C* and 0 for line *B*. Thus, line *C* is raised 1.50 units above line *B*.

The point at which a line intersects the *Y* axis when $X = 0$ is called the *intercept*, and its value is denoted by the letter *a*. Another way of thinking about the intercept is that it is the value of *Y* when *X* is zero.

Linear relationships can differ in the values of their intercepts as well as the values of their slopes. The general form of a linear equation is

$$Y = a + bX \tag{8.3}$$

Stated more formally, the linear function is

$$f(X) = a + bX$$

where a and b are constants representing an intercept and slope and X is a variable. A variable, Y, is described by this function if $Y = f(X)$, that is, $Y = a + bX$. Equation 8.3 is called a *linear equation*.

DETERMINISTIC VERSUS STOCHASTIC MODELS

Any linear relationship can be represented by Equation 8.3. A slope and intercept always describe the linear relationship between two variables. Given values of the slope and intercept, we can substitute scores on X into the linear equation to determine the corresponding scores on Y. For example, the linear equation $Y = 1.50 + 2.00X$ tells us that an individual who works for 2 hours is paid \$5.50 because the Y score associated with an X score of 2 is

$$Y = 1.50 + 2.00X = 1.50 + (2.00)(2) = \$5.50$$

An individual who works for 3 hours is paid $Y = 1.50 + (2.00)(3) = \$7.50$, and an individual who works for 4 hours is paid $Y = 1.50 + (2.00)(4) = \$9.50$.

When one variable is a linear function of another, all the data points on a scatterplot will fall on a straight line. However, rarely in the social sciences will we encounter such situations. When two variables only approximate a linear relationship, we need to add a term to the linear equation to accommodate random disparities from linearity. The term is called a *disturbance* or *error term*, yielding the equation

$$Y = a + bX + e$$

where e is the difference between the observed Y score and the predicted score based on the linear function. The errors are assumed to be random rather than systematic because if the errors were systematic, then some meaningful form of nonlinearity would be suggested and would need to be modeled. It is important to keep in mind that e is an unmeasured variable that reflects the disparity between scores predicted by the model and observed scores.

Formal mathematical models do not include an error term when specified at the theoretical level. In this sense, they are deterministic rather than probabilistic. However, when testing mathematical models empirically, it is common for researchers to include an error term because there usually is some random "noise" that creates disparities from model predictions. A common practice is to identify the function that seems appropriate for predicting and understanding a phenomenon and then, in data-based tests of the model, to add an error term to accommodate the hopefully small but random discrepancies that seem inevitable. A model is better if the discrepancies from predictions are trivial and have no practical consequence.

In the world of mathematical models, you will encounter distinctions between deterministic and probabilistic models. A *deterministic model* is one in which there is no random error operating. The model performs the same way for any given set of conditions. In contrast, a *probabilistic model* is one in which some degree of randomness is present. Probabilistic models are also sometimes referred to as *stochastic models*.

MODEL PARAMETERS

Adjustable Parameters and Parameter Estimation

Mathematical models typically include variables that are measured as well as constants whose values can be derived logically or estimated from data. For example, in the linear function

$$f(X) = a + bX$$

there is a variable, X, and two model constants that need to be specified, the intercept, a, and the slope, b. Constants such as the intercept and the slope are called *adjustable parameters* or *adjustable constants*, because their values can be set by the theorist to different values so as to affect the output of the function. For example, we might state that annual income is a function of the number of years of education, where the function is defined as $f(X) = 1{,}000 + 5{,}000X$. If the number of years of education is 2, then output of the function is $1{,}000 + (5{,}000)(2) = 11{,}000$. By contrast, we might state that the function is $f(X) = 2{,}000 + 4{,}000X$. If the number of years of education is 2, then the output of the function is $2{,}000 + (4{,}000)(2) = 10{,}000$. The slope and intercept are adjustable constants that affect the output value of the function as different values of X are substituted into the function.

When one is unsure what the value of the adjustable constants should be, then strategies can be used to estimate their values empirically based on data. For example, in a linear model where the relationship between Y and X is linear, except for the presence of random noise (i.e., $Y = a + bX + e$), a researcher might obtain data for the values of Y and the values of X for a group of individuals and then use traditional least-squares regression methods to estimate the values of the intercept and slope.

Mathematical models differ in the number of adjustable constants they include and in the number of constants that must be estimated from data. Models with many parameter values that must be estimated are less parsimonious and often present greater challenges for testing than models with fewer estimated parameters. When the value of an adjustable parameter is specified a priori by the theorist and not estimated, it is said to be *fixed*. When the value of the adjustable parameter is estimated from data, it is said to be *estimated*. Thus, you will hear reference to fixed parameters and estimated parameters in math models.

RATES AND CHANGE: DERIVATIVES AND DIFFERENTIATION

Parameters in a mathematical model often are subject to meaningful interpretation. In the linear model, $Y = a + bX$, the slope reflects the predicted change in Y given a 1-unit change in X. It is calculated using Equation 8.1, which we repeat here:

$$b = (Y_2 - Y_1)/(X_2 - X_1)$$

The slope is meaningful because it provides a sense of how much change in Y we can expect, given a change in X. Note that in the linear model, it does not matter where the change occurs on the X continuum. A 1-unit change on X, at the low end of the X continuum, will produce the same amount of change in Y as a 1-unit change in X at the high end of the X continuum. The value of the slope tells us how much change this is.

The slope is, in essence, a rate of change in Y, given a unit change in X. More generally, if we describe the change in Y between any two points as

$$\Delta Y = Y_2 - Y_1$$

and the change in X between those same two points as

$$\Delta X = X_2 - X_1$$

then the rate of change in Y relative to the change in X is the ratio of these

$$\text{Rate of change} = \frac{\Delta Y}{\Delta X} = \frac{Y_2 - Y_1}{X_2 - X_1}$$

which, in this case, is the value of the slope. If $\Delta Y = 4$ and $\Delta X = 2$, then the rate of change of Y relative to a unit change in X is $4/2 = 2$.

The property of equal amounts of change at all points on the X continuum does not apply to nonlinear relationships. Consider the nonlinear relationship between Y and X shown in Figure 8.4. At low values of X, small changes in X result in no change in Y, whereas at high values of X, small changes in X result in large changes in Y. The impact of a 1-unit change in X differs depending on the part of the X continuum at which the change occurs.

When analyzing change, two fundamental concepts from calculus are helpful: derivatives and differentiation. *Derivatives* refer to the concept of instantaneous change, and *differentiation* refers to algebraic methods for calculating the amount of instantaneous change that occurs. Let us explore these concepts in more depth.

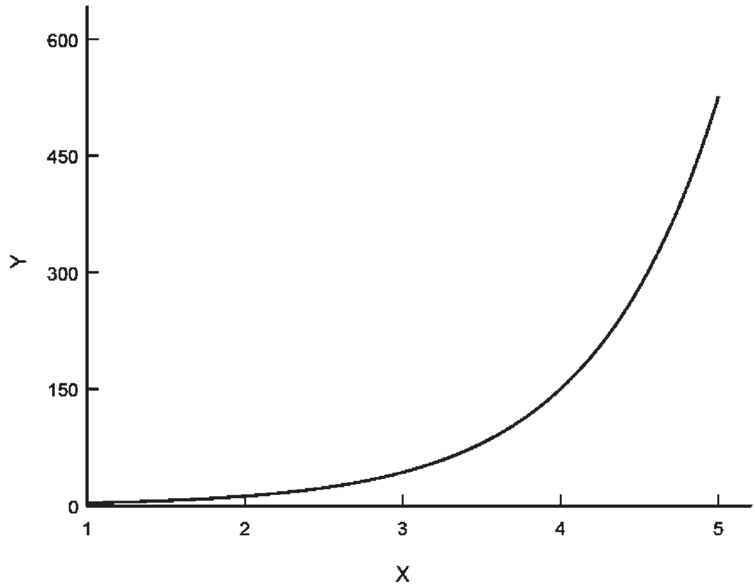


FIGURE 8.4. Nonlinear relationship for derivative example.

Instantaneous Change

Suppose we want to measure the speed of a car driving between two towns, Town A and Town B, that are 120 miles apart. Let Y be the distance traveled by the car. When the car is in Town A and just about to begin its journey, the car has traveled 0 miles, so we set $Y_1 = 0$. When the car reaches Town B, it has traveled 120 miles, so we set $Y_2 = 120$. Now let X be the amount of time the car spends traveling. Before the car leaves Town A, $X_1 =$ 0 hours. Suppose when the car finally reaches Town B, the car has been on the road for 2 hours. This means that $X_2 = 2$ hours. Using the logic from above, the rate of change in Y as a function of X is

$$\text{Rate of change} = \frac{(Y_2 - Y_1)}{(X_2 - X_1)} = \frac{\Delta Y}{\Delta X} = \frac{(120-0)}{(2-0)} = 60 \tag{8.4}$$

or 60 miles per hour. A 1-unit change in time (X, as measured in hours) is associated with a 60-unit change in distance (Y, as measures in miles).

The value of 60 miles per hour represents the average speed of the car during the entire trip. But it is probably the case that the car did not travel at a speed of exactly 60 miles per hour during the entire trip. At times, it probably was driven faster and at other times, slower. Suppose we wanted to know how fast the car was going 15 minutes into the trip. One way of determining this number is to define values for X_1 and Y_1 at 14 minutes and 59 seconds into the trip and then to define X_2 and Y_2 values at 15 minutes and 1 second into trip. We could then apply Equation 8.4 to this more narrowly defined time frame. Although the result would give us a sense of how fast the car was being driven 15 minutes into the trip, it would not tell us how fast the car was being driven at *exactly* 15 minutes into the trip. We want to know at the very instant of 15 minutes into the trip, how fast the car was going, that is, what its rate of change was at that particular instant. It is this concept of instantaneous change to which a derivative refers. The velocity that the car is traveling at an exact point in time maps onto the notion of a derivative.

For a nonlinear relationship such as that in Figure 8.4, it is possible to use differentiation to calculate the instantaneous rate of change in Y at any given value of X. The derivative is the (instantaneous) slope of Y on X at that given point of X. It is analogous to specifying the speed at which a car is being driven at a specific point in time. For some modeling problems, calculating a derivative by the process of differentiation is straightforward. For other problems, it can be quite complex. Methods of differentiation are taught in calculus and need not concern us here. The main point we want to convey is that in many forms of mathematical modeling, rates of change in Y as a function of changes in X are described using the language of derivatives, and it is important that you have a sense of that language.

For linear models, the instantaneous rate of change in Y at some point on the X continuum is the same as the instantaneous rate of change in Y at any other point on the X continuum. By contrast, for the nonlinear relationship in Figure 8.4, the instantaneous rate of change depends on where on the X continuum the change is occurring. In Figure 8.4 the derivative (i.e., instantaneous rate of change) when $X = 1$ is 0.04, whereas when $X = 4$, the derivative is 1.98. We calculated these values using calculus. A common notation for signifying a derivative is dY/dX. A common phrase for describing derivatives is to state "the value of the derivative at $X = 4$ is 1.98." If the derivative has the same value at all points on X (as is the case for a linear relationship), then one refers simply to "the derivative" without specifying the value of X at which the derivative is calculated.

You also may encounter a derivative expressed as a rate of change (ΔY and ΔX), but invoking what is called a limit, perhaps as follows:

$$\lim_{\Delta X \to 0} \frac{\Delta Y}{\Delta X}$$

The left-hand part of this expression contains the abbreviation *lim* (for the word *limit*), and the entire expression describes symbolically the idea of instantaneous change. Specifically, this expression is read as "the change in Y relative to the change in X as the change in X approaches its lower limit of zero" (analogous to the case where we calculated speed at exactly 15 minutes into the trip). The expression is just a way of referring to a derivative in a more formal way.

Second and Third Derivatives

In calculus some functions have higher-order derivatives, such as a second derivative or a third derivative. We will not use second or third derivatives in the mathematical models considered in this chapter, but it will help to have some appreciation for these concepts. As noted above, a derivative refers to a rate of change of one variable (ΔY) relative to the rate of change of another variable (ΔX) in the context of instantaneous change. In our driving example, the first derivative refers to the speed or velocity with which a car is driven at any given point in time. Suppose the car is driving along and the driver decides to speed up. The result of pressing harder on the accelerator is that the car's velocity (i.e., the first derivative) increases. A second derivative in this case refers to the change in the first derivative that occurred at any given instant. It is analogous to what we commonly call acceleration. When you press the accelerator pedal, you "change" your speed. How much your speed changes at a given instant is the idea of a second derivative. If, in turn, your rate of acceleration changes (e.g., you "let off" the pedal and decelerate), then this maps onto the idea of a third derivative.

In a linear function the value of the second derivative is zero, because there is never a change in the value of the first derivative at different points of X. For nonlinear relationships, the value of the second derivative is nonzero at different points on X. When reading about mathematical models, in addition to the concept of first derivatives, you may encounter the concepts of second or third derivatives.

In sum, derivatives are useful concepts for describing rates of change in Y as a function of X. For nonlinear functions, the rate of change in Y will differ depending on where on the X continuum the change is occurring. A derivative is an index of instantaneous change at a given X value. It is a slope, but a special one, namely an "instantaneous" slope. First derivatives are fairly straightforward. Second and third derivatives are a bit more abstract. For those readers familiar with interaction effects in statistics, a second-order derivative is roughly analogous to a two-way interaction and a third order derivative is roughly analogous to a three-way interaction.

DESCRIBING ACCUMULATION: INTEGRALS AND INTEGRATION

When describing mathematical models, many theorists emphasize derivatives, that is, rates of change. There is another concept in calculus that is sometimes emphasized in mathematical models: the *integral*. This concept reflects the amount of something. The process of calculating an integral is called *integration*. To gain a sense of what an integral is, consider the well-known function in statistics of the probability density function for a standard normal distribution. This function, often presented in statistics texts, is the basis for calculating the "area under the curve" in a normal distribution. Figure 8.5 presents a graphical representation of this function, as it often appears in statistics books. The various X values on the horizontal axis are standard scores, with a mean of zero and a standard deviation of 1. One can specify any two points in this distribution, say, a value of 0 and a value of 1, and then calculate the area under the curve between these two points. If one scales the total area under the curve to equal a value of 1.00, then the area under the curve between two X scores is the proportion of the total area that falls between the two points. For example, the area under the curve between the X values of 0 and 1 is 0.3413 (see Figure 8.5). Between the X values of 1 and 2, the area under the curve is 0.1359. Between X values of –1 and 1, the area under the curve is 0.6826.

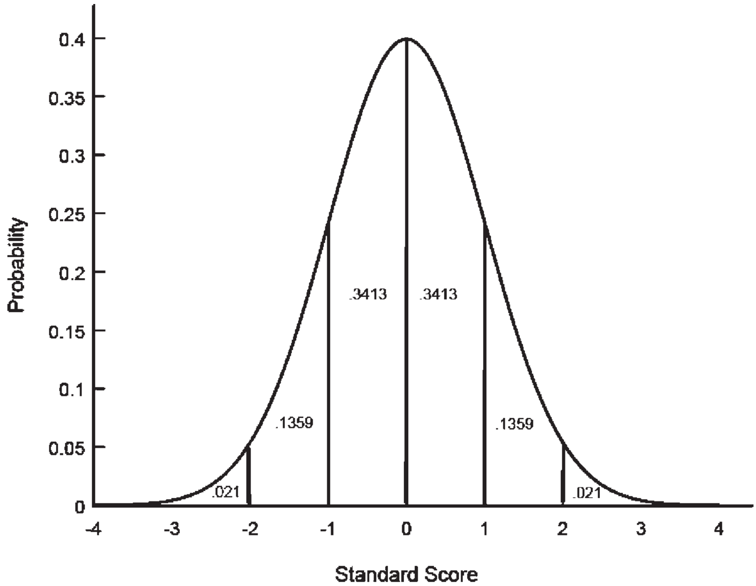


FIGURE 8.5. Area under the curve for a standardized normal distribution.

Graphically, an integral is the area under the curve between two points. The integral for the values 0 and 1 in Figure 8.5 is 0.3413. The integral for the values 1 and 2 is 0.1359. Because it focuses on the area under the curve, one can see that, roughly speaking, an integral refers to "the amount of something."

A common use of integrals in mathematical models is to characterize accumulations, that is, how much of something has accumulated. Many phenomena accumulate. We accumulate money in savings accounts. Frustration accumulates with each stressful event experienced within a short time span. Although the mathematical details of integration are well beyond the scope of this book, when one uses the concept of integrals, one often does so in the context of building models of phenomena that accumulate.

JUST-IDENTIFIED, OVERIDENTIFIED, AND UNDERIDENTIFIED MODELS

Mathematical models vary in their identification status. *Model identification* refers to cases where the values of model parameters must be estimated from data. A *just-identified* model is one for which there is a unique solution (i.e., one and only one solution) for the value of each estimated parameter in the model. Consider an analogy from algebra, where we might be given two equations with two unknowns of the following form:

$$23 = 2X + 3Z$$
$$9 = X + Z$$

For these two equations, there is a unique solution for X and Z: $X = 4$ and $Z = 5$.

An *underidentified* model is one for which there is an infinite number of solutions for one or more of the model parameters. In the equation

$$10 = X + Z$$

there is an infinite number of solutions for X and Z (e.g., $X = 10$ and $Z = 0$ is one solution; $X = 9$ and $Z = 1$ is another solution). Models that have one or more parameters that are underidentified are often problematic.

An *overidentified* model is one for which there is a unique solution for the model parameters, *and* there is more than one feature of the model that can be used to independently estimate the parameter values. Using the algebraic analogy, consider the following three equations:

$$\begin{aligned} 10 &= X + Z \\ 18 &= 2X + Z \\ 12 &= X + 2Z \end{aligned}$$

There are three equations with a total of two unknowns, and any given pair of equations, no matter which pair, can be used independently to solve for the unknowns. In models for which parameter values must be estimated and the function fit is not perfect (i.e., there is an error term such that the model is stochastic), model parameters that have overidentified status are desirable because one can obtain independent estimates of those parameter values.

In sum, when reading math models, you may encounter references to a model as being just-identified, underidentified, or overidentified. Models that are underidentified are unsatisfactory.

METRICS

When developing mathematical models, theorists give careful consideration to the metric on which the variables in the model are measured, especially when nonlinear modeling is involved. This is because the accuracy of a mathematical model and the inferences one makes can be (but are not always) influenced by the metric of the variables. Depending on the variable metric, a theorist might resort to different functions in the mathematical model to describe the relationships between variables. For example, for the variable time, the model form and parameters introduced into the model might vary depending on whether time is measured in milliseconds, seconds, days, weeks, or years. The nature of metrics poses difficulty for some constructs in the social sciences because the metric used to measure them is arbitrary. For example, when a researcher uses a 10-item agree–disagree scale to measure peoples' attitudes toward religion, the metric might be scored from –5 to +5, or from 0 to 9, or from 1 to 10. In some mathematical models, the choice of scoring matters a great deal, so an arbitrary metric can create modeling difficulties.

TYPES OF NONLINEARITY

Thus far we have considered a simple mathematical model—the linear model—to introduce several concepts in mathematical modeling. The linear model has two adjustable parameters, a slope and an intercept, that typically are estimated rather than fixed by the theorist. In this section we introduce other functions that are nonlinear in form and that can be used in mathematical models. There are many classes of functions, and we cannot begin to describe them all. Here we focus on describing five major classes of functions (the linear function makes six classes). The idea is to give you a sense of some of the nonlinear functions that can be used to build a math model. After presenting the functions, we describe strategies for modifying and combining them to build even more intricate mathematical representations. As we describe the different functions and the modifications to them that can be made, you will see the wide range of tools available to a math modeler when characterizing relationships between variables.[2]

To describe functions, we often use three concepts: (1) concavity, (2) proportionality, and (3) scaling constants. *Concavity* refers to whether the rate of change on a curve (the first derivative) is increasing or decreasing. A curve that is *concave upward* has an increasing first derivative, and a curve that is *concave downward* has a decreasing first derivative. In terms of proportionality, two variables are proportional to one another when one variable is a multiple of the other. More formally, Y is proportional to X if $Y = cX$, where c is a constant. The value c is called the *constant of proportionality*. For proportionality, doubling X doubles Y, tripling X triples Y, and halving X halves Y. Two variables are said to be *inversely proportional* when there is some constant c for which $Y = c/X$. In this case, doubling X halves Y, tripling X cuts Y by one-third, and halving X doubles Y. *Scaling constants* refer to adjustable parameters in a model that have no substantive meaning but are included to shift a variable from one metric to another. For example, to change meters to centimeters, we multiply the meters by the constant 100, which shifts the metric of length to centimeters. As we describe different functions below, we occasionally do so in terms of the concepts of concavity, proportionality, or scaling constants.

Logarithmic Functions

The logarithmic function (often referred to as the *log function*) has the general form $f(X) = \log_a(X)$, where a is a constant indicating the base of the logarithm. Logs can be calculated for different bases, such as the base 10, the base 2, or the base 8.

[2]Our discussion of functions and the graphic representations of them draws on concepts described by W. Mueller (see *www.wmueller.com/precalculus/index.html*).

The log base 10 of the number 100 is written as $\log_{10}(100)$, where the subscript is the base and the number in parentheses is the number for which you are calculating the log. If n stands for the number for which you are calculating the log, and a is the base of the log, then the log is the solution for b in the equation $n = a^b$. For $\log_{10}(100)$, we solve for b in the equation $100 = 10^b$, so the log base 10 of the number 100 is 2 (because $10^2 = 100$). The value of $\log_5(25)$ is 2 because $5^2 = 25$. Sometimes you will encounter a log expression with no base, such as log(1,000). When this happens, the log is assumed to have a base of 10. So log(1,000) = 3 (because $10^3 = 1{,}000$).

There is a special logarithm, called the natural log, that uses a constant called e as its base. The number e appears in many mathematical theories. Its value is approximately 2.71828. The number e was studied in depth by Leonhard Euler in the 1720s, although it was first studied by John Napier, the inventor of logarithms, in 1614. It has some remarkable mathematical properties (which we will not elaborate on here) and is referred to as Napier's constant. The natural log of a number is signified by the expression $\ln(n)$. For example, the natural log of 10 is signified by ln(10). It equals approximately $\log_{2.71828}(10) = 2.302585$.

Figure 8.6 presents sample graphs of log functions. When expressing the relationship between two variables, rather than using a linear function, one might use a log function. Log functions are sometimes used to model growth or change when the change is rapid at first and then slows down to a gradual and eventually almost nonexistent pace (see Figure 8.6a).

Log functions share many common features: (1) The logarithm is undefined for negative values of X (where X is the number for which you are calculating the log); (2) the value of the log can be positive or negative; (3) as the value of X approaches zero, the value of the log approaches negative infinity; (4) when $X = 1$, the value of the log is 0; and (5) as X approaches infinity, the log of X also approaches infinity. For the function $\log_a(X)$, the function output increases with increasing X if $a > 1$ and decreases with increasing X if a is between 0 and 1.

Exponential Functions

The exponential function has the general form $f(X) = a^X$. The function yields output that increases in value with increasing X if $a > 1$ and decreases in value with increasing X if a is between 0 and 1. Figure 8.7 presents examples of common exponential curves. These curves are often used to refer to growth, such as when people say a population is "growing exponentially." With exponential growth or change, the larger a population gets, the faster it grows. For decreasing exponential growth or change, the smaller the population gets, the more slowly it decreases in size. As it turns out, exponential functions are simply the inverse of log functions, so the two functions mirror image each other's properties. For exponential functions, if a is between –1 and 0, then the output value is a damped oscillation as X increases, and if a is < -1, it is an undamped oscillation as X increases (see the later section on trigonometric functions for a discussion of oscillation).

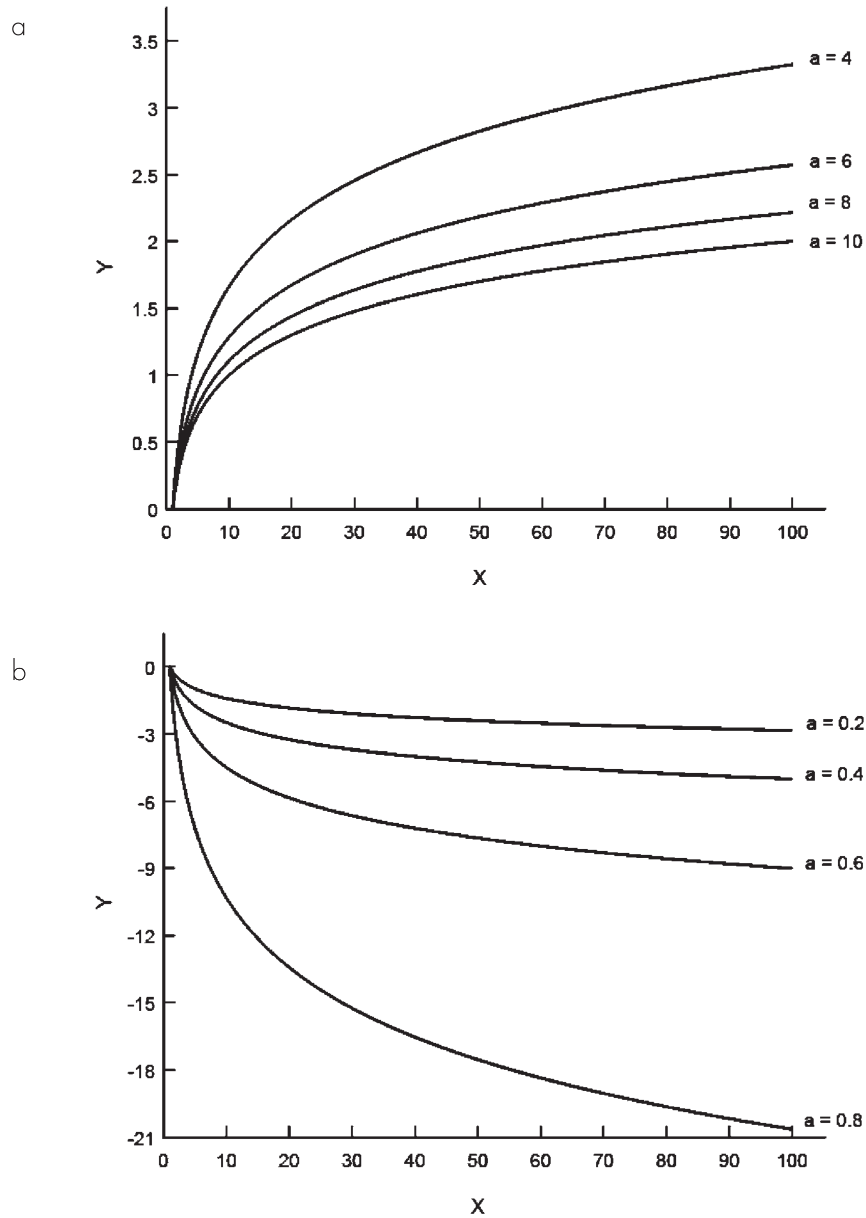


FIGURE 8.6. Graphs of log functions for $\log_a(X)$ with X ranging from 1 to 100. (a) $a > 1$; (b) $0 < a < 1$.

Social scientists often modify the exponential function to create functions that reflect growth or change with certain properties. For example, using the fact that any number raised to the power of 0 is equal to 1, the following equation can be used to describe exponential growth over time

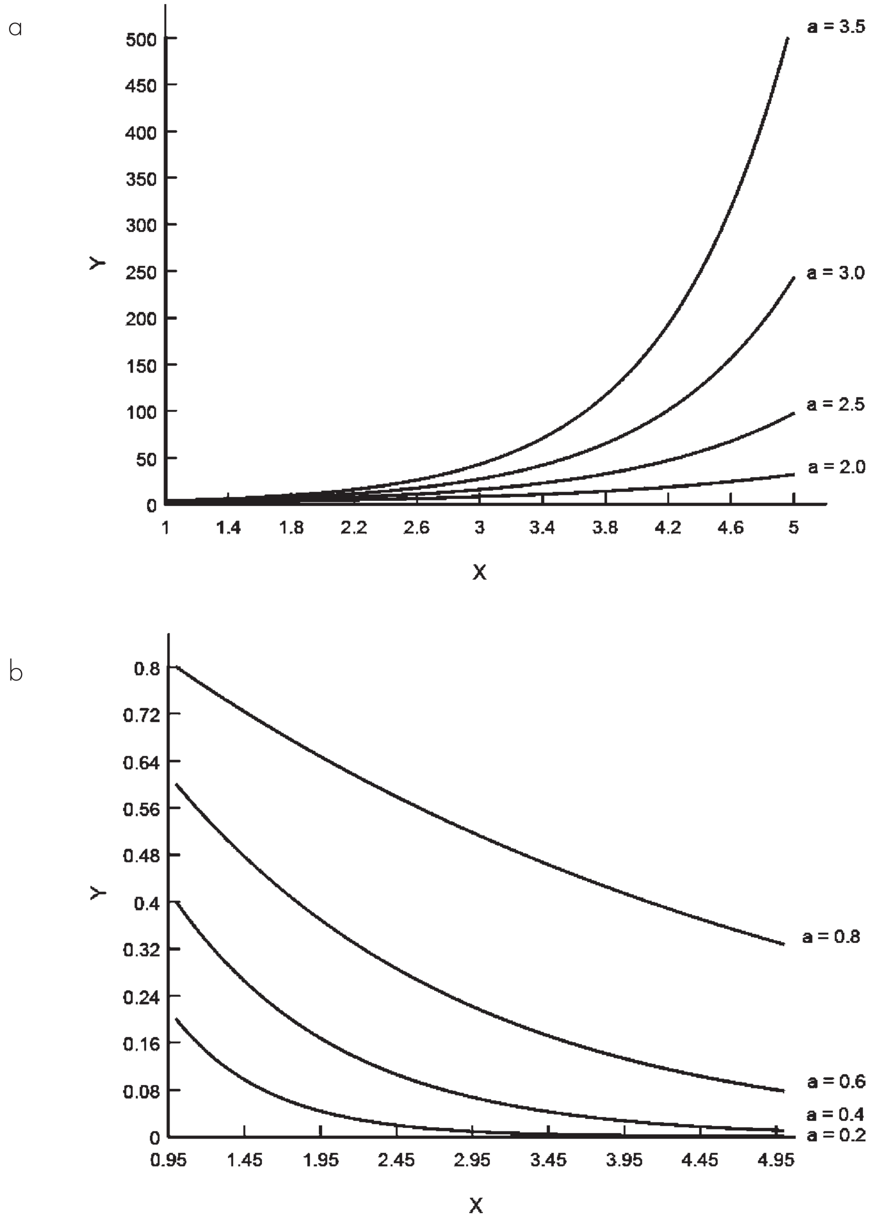


FIGURE 8.7. Graphs of exponential functions for a^X with X ranging from 1 to 5. (a) $a > 1$; (b) $0 < a < 1$.

$$Y = s_0 e^{(kX)}$$

where Y is the population size at a given point in time, X is the duration in time since a predetermined start time, and s_0, e, and k are constants. In this case, e is Napier's constant. The value of s_0 is fixed at a value equal to the population size at the predetermined start time. Note that when $X = 0$, the population size will equal the population size at the start time (because any number raised to the power of zero is 1.0). For this function, Y increases geometrically with a doubling time equal to $0.6932/k$. A graph illustrating this function appears in Figure 8.8, where the starting size of a population is $s_0 = 5{,}000$, where $k = 0.333$ (yielding a doubling time of just over 2 years), and where X ranges from 0 to 5 years. When expressing the relationship between two variables, rather than using a linear or log function, one might use an exponential function, such as that illustrated in Figure 8.8.

Power Functions

Power functions have the general form $f(X) = X^a$ where a is an adjustable constant. For positive values of X greater than 1, when $a > 1$, the curve will be concave upward, and when a is between 0 and 1, the curve will be concave downward.

Power functions often have a similar shape to exponential and logarithmic functions, with the differences between the functions sometimes being subtle. When the difference is small, it does not matter which function is used to create the model. But differences can exist. Exponential functions increase by multiples over constant input intervals. Logarithms increase by constant intervals over input multiples. Power functions do not follow either of these patterns. Power curves eventually outgrow a logarithm and undergrow an exponential as X increases. A practical example of the function differences is the modeling of the spread of HIV, the virus that causes AIDS. During the

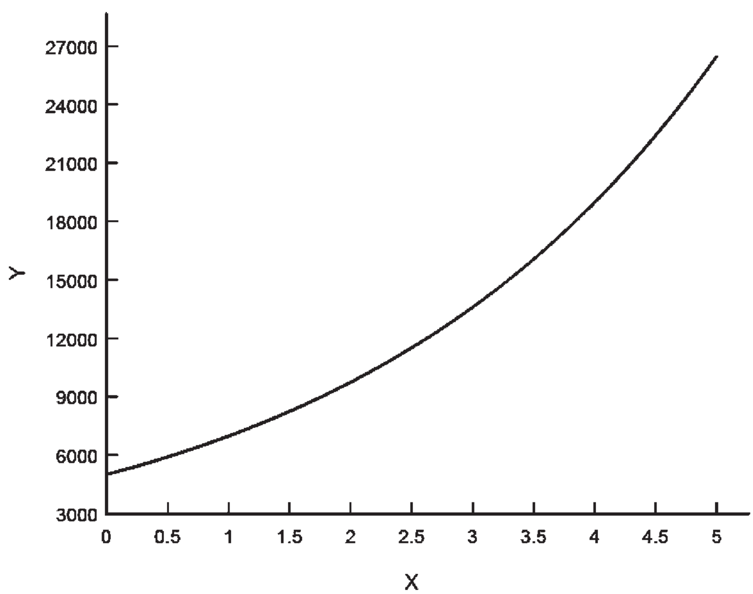


FIGURE 8.8. Exponential growth for $Y = s_0\, e^{(kX)}$.

early stages of the epidemic, it was thought that the number of HIV cases was growing exponentially, but in later analyses, the function was found to be better mapped by a power function. An exponential model yielded overestimates of the number of cases forecast in future years, which in turn led to overestimates of the required resources to deal with the epidemic (e.g., hospital space, medications; see Mueller, 2006).

Figure 8.9 presents some examples of power functions, and Figure 8.10 plots a power function and an exponential function on the same graph to illustrate some of these properties. When expressing the relationship between two variables, rather than using a linear, log, or exponential function, one might use a power function instead.

Polynomial Functions

Polynomial functions are simply the sum of power functions. The general form of a polynomial function is

$$f(X) = a + bX^1 + cX^2 + dX^3 + \ldots$$

where X continues to be raised to the next highest integer value, and each term has a potentially unique adjustable constant. Polynomials can model data with many "wiggles and turns," but the more wiggles and turns there are, the greater the number of power terms that are required to model it. Notice that when only a single term for X is used with a power of 1, the polynomial model reduces to a linear model. The adjustable constant a is typically viewed as a scaling constant. Adding one term to the linear model (i.e., adding the term cX^2) allows the model to accommodate a curve with one bend. A polynomial model with three terms ($a + bX^1 + cX^2 + dX^3$) will accommodate a curve with two bends. A polynomial model with four terms will accommodate a curve with three bends. In general, to accommodate k bends, you need $k + 1$ terms.

The most popular polynomial functions in the social sciences are the quadratic and cubic functions. They are defined as

$$\text{Quadratic: } f(X) = a + bX + cX^2$$
$$\text{Cubic: } f(X) = a + bX + cX^2 + dX^3$$

Figures 8.11 and 8.12 provide an example of each type of curve, and Figure 8.13 provides an example of a polynomial function with eight terms. The quadratic model is effective for modeling U-shaped and inverted-U-shaped relationships as well as J-shaped and inverted-J-shaped relationships. The cubic function is effective for modeling S-shaped curves. In Figure 8.12b we manipulated the scaling constant, a, to create separation between curves so that you can better see the trends. Figure 8.13 illustrates how diverse a "curve" that large polynomials can create. When expressing the relationship between two variables, rather than using a linear, log, exponential, or power function, one might use a polynomial function instead.

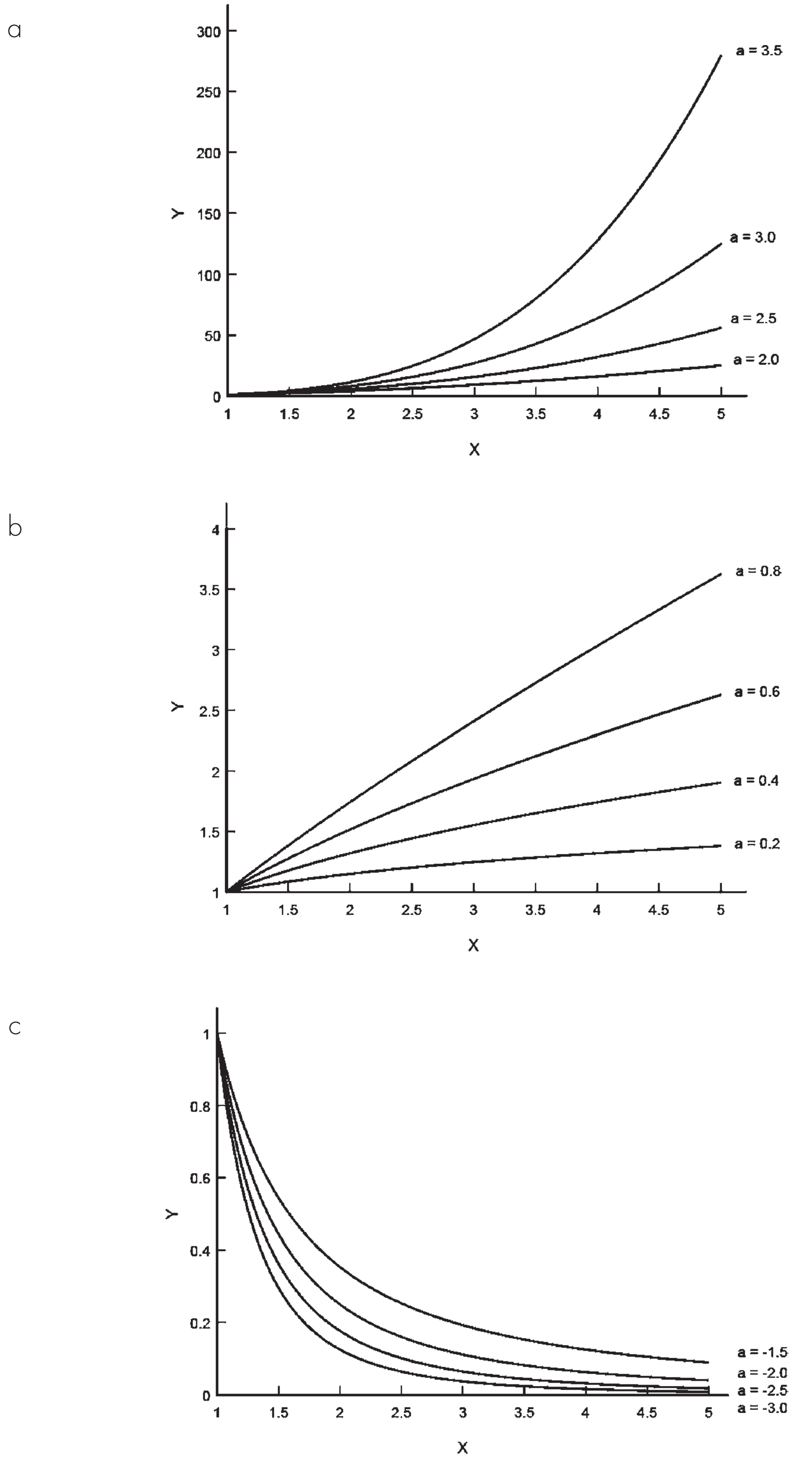


FIGURE 8.9. Graphs of power functions for X^a with X ranging from 1 to 5. (a) $a > 1$; (b) $0 < a < 1$; (c) $a < 0$.

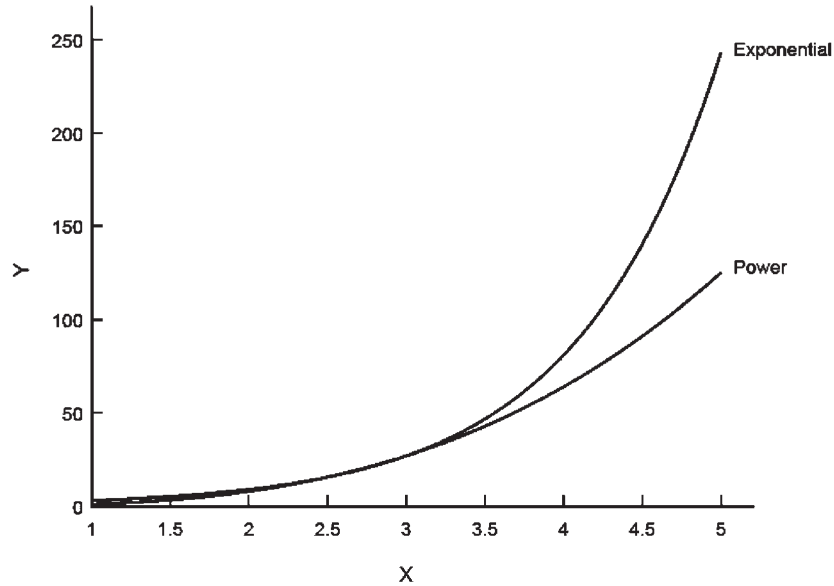


FIGURE 8.10. Example of power and exponential functions.

Trigonomic Functions

Trigonometric functions are typically used to model cyclical phenomena. The two most common functions are the sine function and the cosine function, which have the form $f(X) = \sin(aX)$ and $f(X) = \cos(aX)$, where a is a constant, *sin* is the sine, and *cos* is the cosine. The sine and the cosine functions repeat the values of their outputs at regular intervals as X increases. Simple transformations of the sine and cosine functions can reproduce many forms of periodic behavior. For example, some people have suggested that rhythmic cycles, called biorhythms, reflect active and passive phases in the physical aspects of everyday life. The phases of biorhythms are modeled using a sine function of the form $f(X) = \sin(.224*X)$, where X is the number of days since a baseline index is taken. Output values range from 1 to –1, with positive values indicating increasingly high energy and negative values indicating increasingly low energy. Figure 8.14 plots the output values for 120 days, starting at day 0. As noted earlier, certain types of cyclical phenomena also can be modeled using exponential functions with negative values of a in the expression $f(X) = a^X$.

Choosing a Function

In sum, there are a wide range of functions available to the math modeler for describing the relationship between variables, including linear functions, logarithmic functions, exponential functions, power functions, polynomial functions, and trigonometric functions, to name a few. We have only scratched the surface of the many strategies a

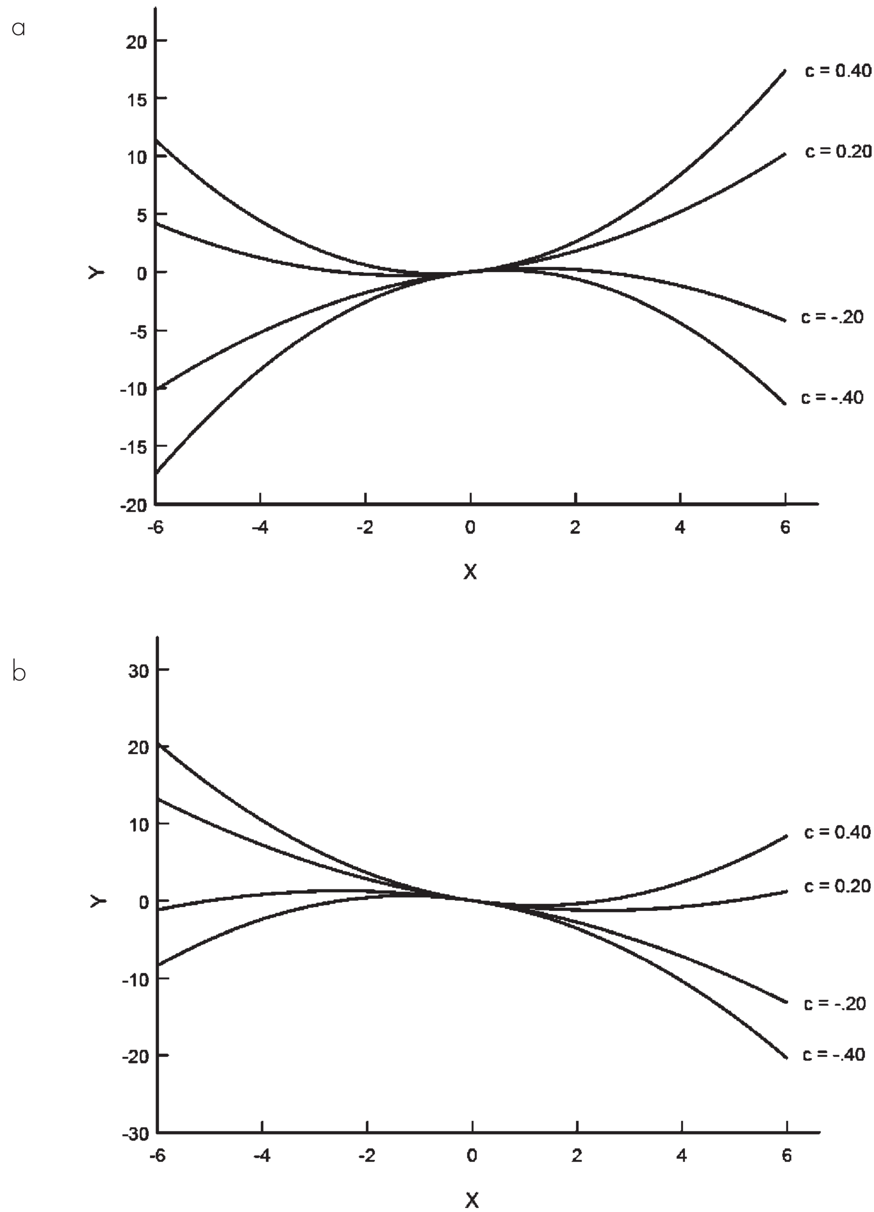


FIGURE 8.11. Quadratic functions. (a) Function $a + bX + cX^2$, with $a = 0$, $b = .5$; (b) function $a + bX + cX^2$, with $a = 0$, $b = -1$.

mathematical modeler can use. As you become familiar with functions and the curves they imply, you should be able to make informed choices about modeling relationships between variables. Mathematical modelers sometimes select functions for their models a priori, based on logic, and other times they make decisions about appropriate model a

b

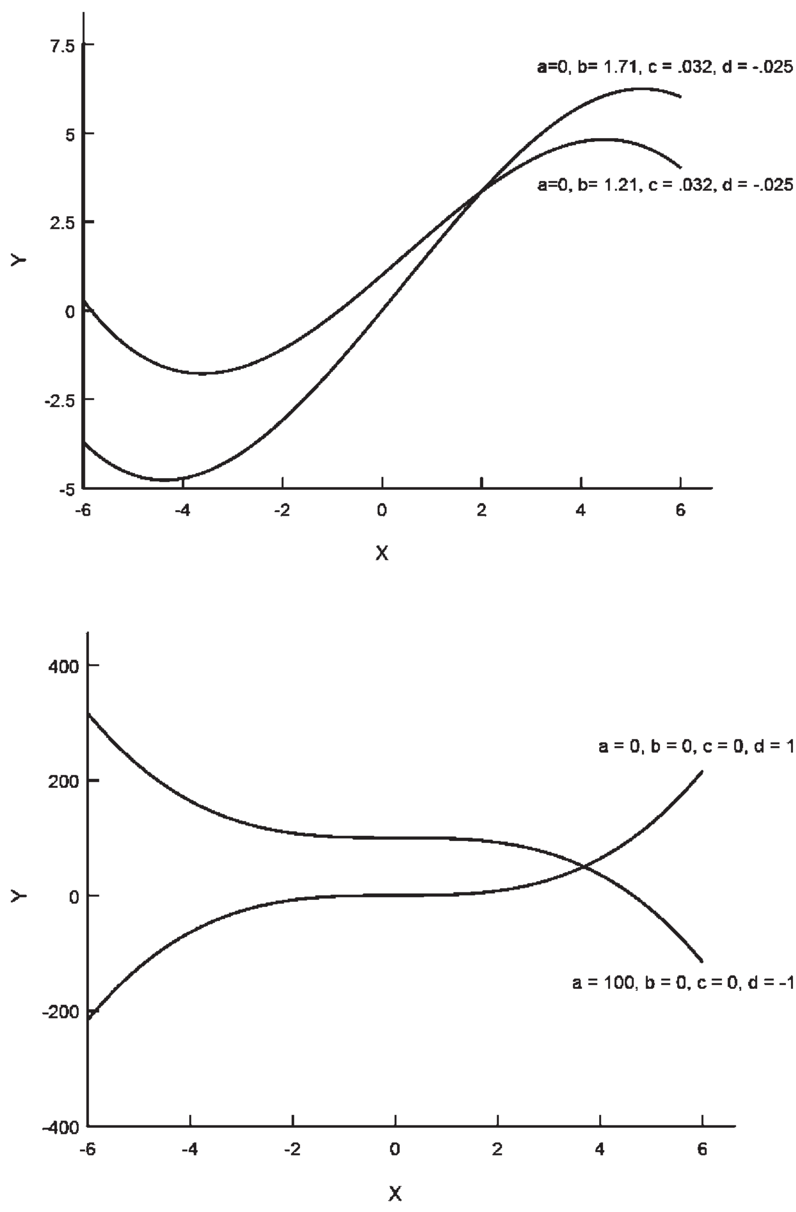


FIGURE 8.12. Cubic functions. (a) Function for $a + bX + cX^2 + dX^3$; (b) additional functions for $a + bX + cX^2 + dX^3$.

functions after collecting data and scrutinizing scatterplots. In the latter case, the model chosen and the values of the adjustable parameters are still subjected to future empirical tests, even though preliminary data are used to gain perspectives on appropriate functional forms. You can gain perspectives on the curves implied by different functions by creating hypothetical data and applying the different functions to them. We provide information on how to do this using the statistical package SPSS in Appendix 8A to this chapter, and also provide information about other graphics software.

FUNCTIONS FOR CATEGORICAL VARIABLES

Thus far we have considered only functions involving quantitative variables, but functions also can be specified for categorical variables. Consider as a simple example the relationship between whether or not someone uses an umbrella as a function of whether or not it is raining. The relationship between these two categorical variables is expressed as follows

$$f(x) = \begin{cases} \text{umbrella}, & \text{if } x = \text{raining} \\ \text{no umbrella}, & \text{if } x = \text{not raining} \end{cases}$$

where one uses an umbrella if it is raining and one does not use an umbrella if it is not raining.

Sometimes mathematical modelers create quantitative representations of categorical variables and then analyze the quantitative translations using the quantitative functions described earlier. For example, one could specify a mathematical function relating the probability of carrying an umbrella to the probability of it raining, with both variables

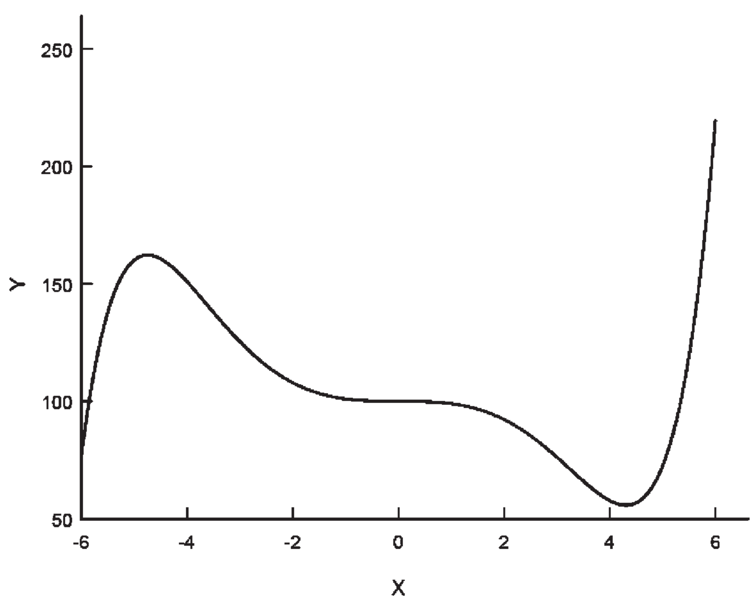


FIGURE 8.13. Polynomial function with seven terms.

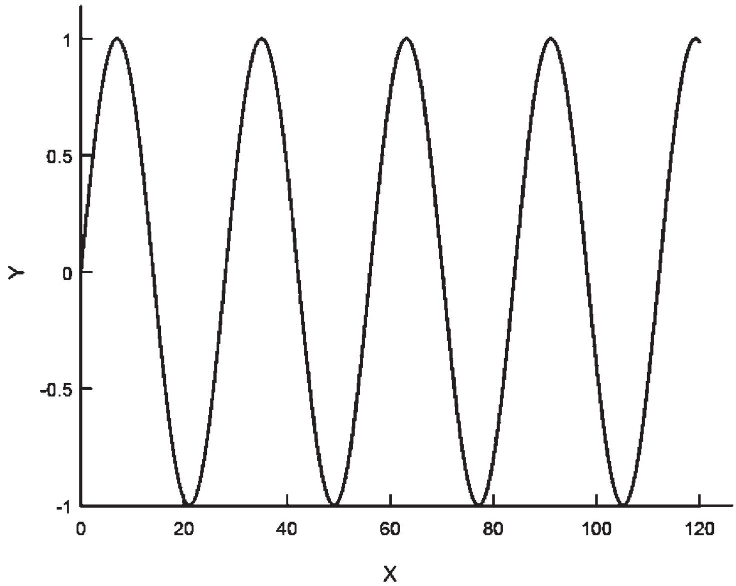


FIGURE 8.14. Sine function.

differing on a probability continuum of 0 to 1.0. The function might then be expressed as an exponential function, as in Figure 8.7a.

In some cases, functions involving categorical and quantitative variables are stated in terms of a table of values rather than symbolically. For example, suppose we specify whether someone is a Democrat or Republican as a function of scores on a 7-point index (e.g., response to a rating scale) of how conservative or liberal he or she is. The scale consists of integers ranging from –3 to +3, with increasingly negative scores signifying more conservativeness, increasingly positive scores signifying more liberalness, and the score of zero representing a neutral point. The function $Y = f(X)$ might be stated as

X	*Y*
–3	Republican
–2	Republican
–1	Republican
0	Democrat
1	Democrat
2	Democrat
3	Democrat

In this representation, the person is said to be a Republican if he or she has a value of –1, –2, or –3. Otherwise, the person is a Democrat.

Another approach to representing a function with a categorical variable is to use a graph. For example, the liberal–conservative and party identification function might be expressed as in Figure 8.15.

ADVANCED TOPICS: MANIPULATING AND COMBINING FUNCTIONS

One creative aspect of mathematical modeling is deriving new functions from old functions so as to create models that are better suited to describing the relationship between variables. We saw hints of this for polynomial functions (which combine power functions). Another class of functions, which we did not discuss, divides one polynomial function by a second polynomial function rather than summing polynomials. These are called *rational functions*. We provide illustrations of manipulating and combining functions here to show the flexibility available to the math modeler.

Function Transformations

One way of modifying functions is to add adjustable parameters to them. Given a function $f(x)$, one can add or subtract an adjustable parameter, a, after the rule described by $f(X)$ is applied: that is, $f(X) \pm a$. This has the effect of shifting the output values upward (in the case of addition) or downward (in the case of subtraction). These transformations are called *vertical shifts*. The output of a function also can be multiplied by the parameter a after the rule described by $f(X)$ is applied, that is, $a \times f(X)$. This transformation vertically stretches (when $a > 1$) or squeezes (when $a < 1$) the graph of the function. Such transformations are called *vertical stretches* or *vertical crunches*. Another possibility is to add or subtract a from $f(X)$ before the rule described by $f(X)$ is applied: that is, $f(X + a)$ or $f(X - a)$. These transformations typically move the graph of the function left when adding a positive value of a or right when subtracting a positive value of a. Such transformations are called *horizontal shifts*. Finally, one can multiply X before the rule described by $f(X)$ is applied; that is, $f(aX)$. These transformations horizontally stretch (when $a < 1$) or squeeze (when $a > 1$) the graph of the function. Such transformations are

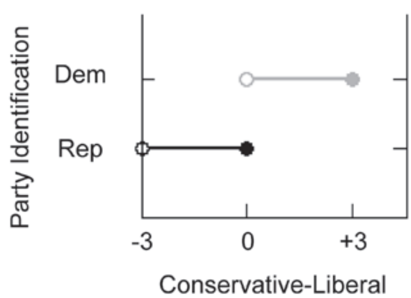


FIGURE 8.15. Graphical representation of a function with a qualitative variable.

called *horizontal stretches* or *horizontal crunches*. Coupled with the possibility of forming inverses for many functions, mathematical modelers have considerable flexibility in manipulating traditional functions with the use of vertical shifts, horizontal shifts, vertical stretches, vertical crunches, horizontal stretches, and horizontal crunches. If you begin your modeling efforts with a traditional function that is approximately correct in form, then transformations such as the above allow you to fine-tune the form of the curve to your problem. An example of this is the classic bounded exponential model, which we now consider.

Recall that the exponential function is $f(X) = a^X$. A simple set of modifications to this function produces what is called a *bounded exponential model*. This has the form

$$Y = a + (b - ce^{-X})$$

where a, b, and c are adjustable constants and e is Napier's constant. The term ce^{-X} is essentially an exponential function where $a = e$ and the exponent is multiplied by an adjustable constant, c. This creates a decaying exponential curve, which is then subtracted from a fixed upper bound or limit reflected by the value of b. As the decaying exponential dies out, the difference from b rises up to the bound. The parameter a is a scaling constant. This kind of function models growth that is limited by some fixed capacity. Figure 8.16 presents an example of this curve, as well as a traditional exponential curve.

Combining Functions

Another strategy that math modelers use is to combine functions. A popular function in the social sciences is the logistic function. It has the general form $f(X) = c/(1 + ae^{-bX})$ where a, b, and c are adjustable constants and e is Napier's constant. A logistic function is a combination of the exponential growth and bounded exponential growth functions that were illustrated in Figure 8.16. In the logistic function, exponential growth occurs when the function outputs for X are small in value. However, this turns into bounded exponential growth as the function outputs approach their upper bound. A logistic function is plotted in Figure 8.17. Note the shapes of the curve to the right and left of the broken line in Figure 8.17 and compare these with the curve shapes in Figure 8.16. The result of combining the exponential growth and the bounded exponential growth functions is an S-shaped curve. The logistic function is a special case of a broader function known as the *sigmoid function*, which generates curves having an S shape.

Combining multiple functions using processes such as those described for the logistic function is another tool available to math modelers. It is not uncommon for a theorist to break the overall relationship into a series of smaller component segments, specify a function to reflect each segment, and then assemble the component functions into a larger whole in one way or another.

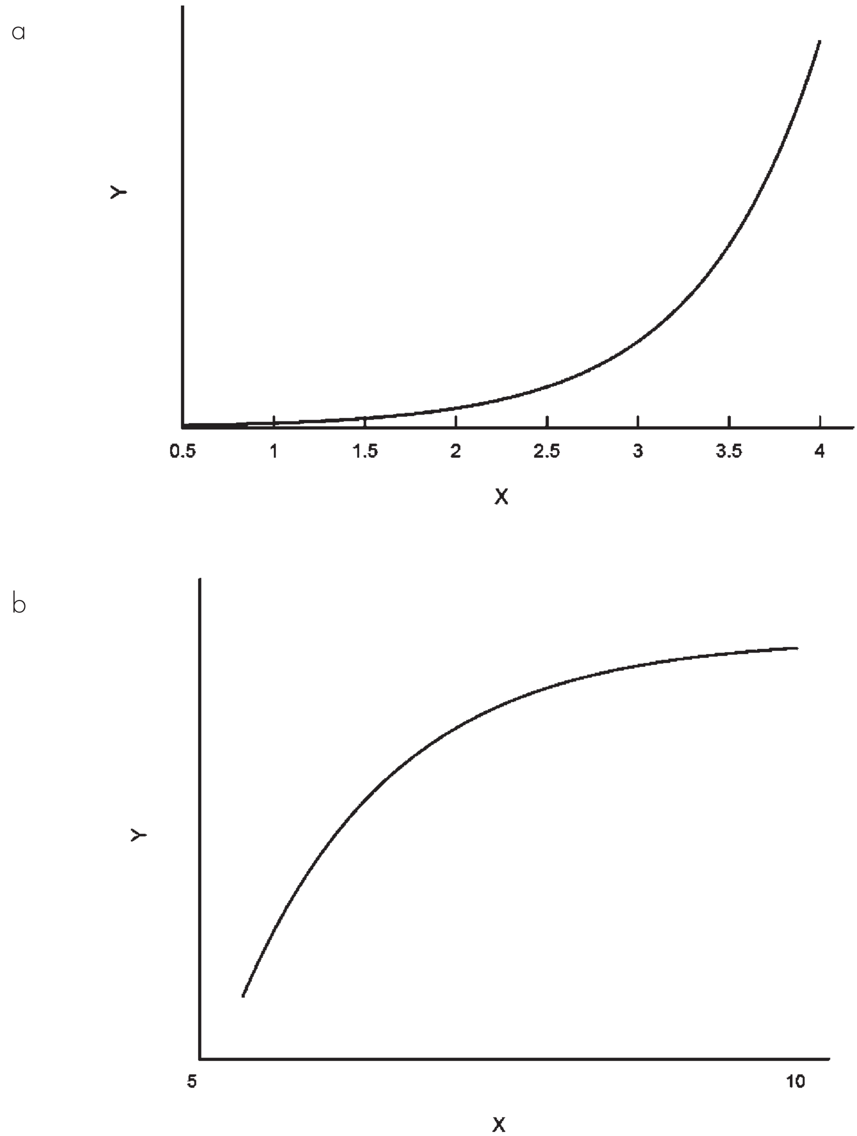


FIGURE 8.16. Exponential and bounded exponential model. (a) Exponential function; (b) bounded exponential function.

In sum, functions can be manipulated with adjustable constants in a variety of ways and subjected to vertical and horizontal stretching and crunching. Functions also can be combined to form even more complex functions (as in the case of the logistic function), and both quantitative and qualitative variables can be modeled. Traditional mathematical modeling opens up a wide range of tools for describing relationships for the theorist to consider.

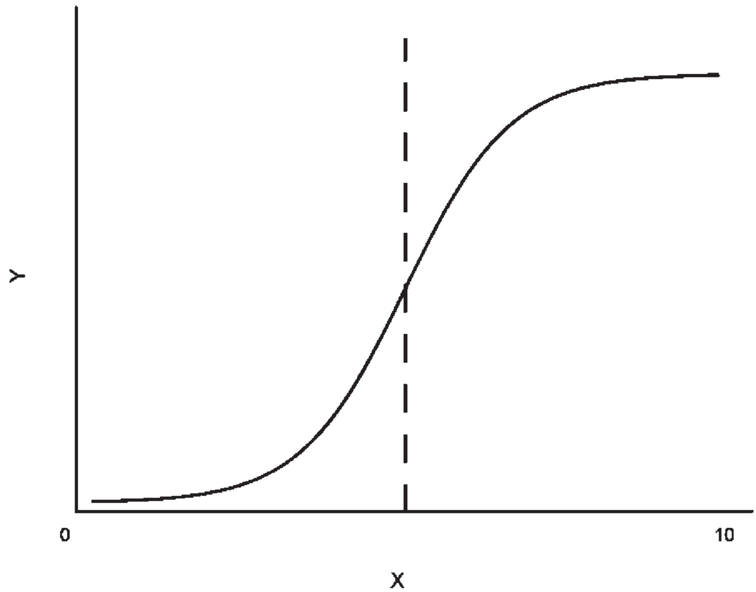


FIGURE 8.17. Logistic function.

MULTIPLE VARIABLE FUNCTIONS

All of the functions we have described use a single input variable. However, functions can involve more than one input variable, and the multiple variables can be combined in a wide variety of ways to yield output. For example, the traditional linear function for a single variable can be extended to include multiple variables (e.g., X and Z) using the following functional form

$$f(X, Z) = a + bX + cZ$$

where a, b, and c are adjustable constants. As another example, a multiplicative function might take the form

$$f(X, Z) = a + bXZ \tag{8.5}$$

where a and b are adjustable constants. Multiplicative models often are used to represent moderated relationships between quantitative variables, as discussed in Chapter 6 (see Jaccard & Turrisi, 2003).

Another example of a multiple variable function that we will make use of later is an averaging function. It takes the general form

$$f(X, Z) = [a/(a + b)]X + [b/(a + b)]Z$$

where a and b are adjustable constants. This model represents function output as a weighted average of X and Z. To see that the function captures a simple arithmetic average, set the values of a and b to 1. This produces

$$\begin{aligned} f(X, Z) &= [1/(1 + 1)]X + [1/(1 + 1)]Z \\ &= (1/2)X + (1/2)Z \\ &= (X + Z)/2 \end{aligned}$$

By allowing a and b to take on nonequal values (e.g., $a = 1$ and $b = 4$), one obtains a "weighted" average rather than a simple arithmetic average, such that the Z will contribute more to the average than X. We will take advantage of this property later when we apply math modeling to a substantive area.

PHASES IN BUILDING A MATHEMATICAL MODEL

Math modelers typically use four phases to construct a mathematical model. First, the modeler identifies the variables that will be included in the model and identifies the metrics on which the variables are measured. Textbooks on mathematical modeling tend to view the variables and metrics as givens and devote little attention to how the variables and metrics are selected. Of course, this is a nontrivial issue, and how one chooses the variables to include is the subject of much of this book. Second, the modeler thinks carefully about the variables, the metrics, and the relationships between the variables, and poses a few candidate functions that might capture the underlying dynamics. He or she might think about the implications of the functions and what predictions to make at both moderate and extreme input values. Eventually, a working function is settled upon, typically a function that includes several adjustable constants. Sometimes the values of the adjustable constants are logically determined, and the modeler fixes the constants at those values. More often than not, the values of the adjustable constants are estimated from data. Third, the modeler collects empirical data, estimates values of the adjustable constants from the data if necessary, and examines the degree of fit between the output values of the function and the values observed in the real world. At this point, if performance of the model is unsatisfactory, a new function might be tried or the original function might be modified to accommodate the disparities. Fourth, given revisions of the function, the model is applied to a new set of data to determine how well the revised model performs. If the model does a good job of reproducing observations in the real world and if the model makes conceptual sense, it will be selected as the model of choice.

This, of course, is an oversimplification of the process that unfolds in building math models, and there are many variants of it that depend on the parameters of the task at hand. Our main point is that building math models is usually a dynamic process that involves much more than simply specifying a function.

AN EXAMPLE USING PERFORMANCE, ABILITY, AND MOTIVATION

Educational researchers have long argued that performance in school is a function of two factors: a student's motivation to perform well and his or her ability to perform well. This relationship is often expressed in the form of a multiplicative model, as follows:

$$\text{Performance} = \text{Ability} \times \text{Motivation} \tag{8.6}$$

The basic idea is that if a student lacks the cognitive skills and capacity to learn, then it does not matter how motivated he or she is; school performance will be poor. Similarly, a student can have very high levels of cognitive skills and the ability to learn, but if the motivation to work and attend to the tasks that school demands is low, then performance will be poor. The multiplicative relationship reflects this dynamic because, for example, if motivation is zero, then it does not matter what a person's score on ability is—his or her performance will always equal zero. Similarly, if ability has a score of zero, it does not matter what a person's motivation score is—his or her performance will always equal zero. Although this makes intuitive sense, the dynamics might be different from those implied by Equation 8.6, as we will now illustrate.

Our first step is to specify the metrics of the variables involved, since they do not have natural metrics. Performance in school might be indexed for individuals using the familiar grade-point average metric that ranges from 1.0 (all F's) to 4.0 (all A's), with decimals rounded to the nearest tenth (e.g., 2.1, 3.5). Ability might be indexed using a standard intelligence test that has a mean of 100 and a standard deviation of 15. Motivation might be indexed using a 10-item scale that asks students to agree or disagree with statements such as "I try hard in school" and "Doing my best in school is very important to me." A 5-point agree–disagree rating scale (1 = strongly disagree, 2 = moderately disagree, 3 = neither agree nor disagree, 4 = moderately agree, and 5 = strongly agree) provides the range of possible responses. The responses to each item are summed to yield an overall score from 10 to 50, with higher scores indicating higher levels of motivation.

Note that neither of these metrics takes on a value of zero. Hence, the dynamic of having "zero" ability or "zero" motivation discussed above cannot occur. Indeed, one might question whether there is such a thing as "zero" intelligence (i.e., a complete absence of intelligence). Is a psychological zero point on this dimension even possible? Suppose we decide that although a complete absence of intelligence is not theoretically plausible, a complete absence of motivation to do well in school is plausible. One way of creating a motivation metric with a zero point is to subtract a score of 10 from the original motivation metric. Before this operation, the motivation metric ranged from 10 to 50. By subtracting 10 from the metric, it now ranges from 0 to 40, which includes a zero point.

However, there is a problem with this strategy. Just because we can mathematically create a zero score on the motivation scale by subtracting 10 from it, this does not mean that the score of zero on the transformed scale reflects a complete absence of motivation on the underlying dimension of motivation. What evidence do we have that this is indeed the case? Perhaps a score of zero on the new metric actually reflects a somewhat low level of motivation but not a complete absence of it. The issue of mapping scores on a metric onto their location on the underlying dimension they represent is complex, and consideration of how to accomplish this is beyond the scope of this book. We will work with the original metric of 10–50 and not make explicit assumptions about where on the underlying motivation dimension these scores locate individuals. We suspect that, based on the content of the items, students who score near 50 are very highly motivated to perform well, and students who score near 10 are very low in (but not completely devoid of) motivation to perform well. But a separate research program is required to establish such assertions (Blanton & Jaccard, 2006a).

Suppose that a student has a score of 100 on the IQ test and a score of 30 on the motivation test. Using Equation 8.6, multiplying the ability score by the motivation score, we obtain 100 × 30 = 3,000, and we would predict a GPA of 3,000! Of course, this is impossible because a student's GPA can range only from 1.0 to 4.0. We need to introduce one or more adjustable constants to Equation 8.6 to accommodate the metric differences and to make it so that a predicted GPA score falls within the 1.0–4.0 range. For example, if we let *P* stand for performance, *A* for ability, and *M* for motivation, then we can allow for the subtraction of a constant from the product to make an adjustment in metric differences, modifying Equation 8.6 as follows

$$P = (A)(M) + a$$

where a is an adjustable constant whose value is estimated from data. Note, for example, if $a = -2{,}997$, then this is the same as subtracting 2,997 from the product of *A* and *M*. But perhaps subtracting a constant is not enough to account for the metric differences. For example, a score of 120 on the IQ test coupled with a score of 50 on the motivation test would yield a product value of 6,000, and subtracting a value of 2,997 from it would still produce a nonsensical GPA. A second scalar adjustment we might use is to multiply the product term by a fractional adjustable constant, which yields the general equation

$$P = b(A)(M) + a$$

where b is a second adjustable constant (in this case, b would be a fraction) designed to deal further with the metric differences. Its value also is estimated from data. The terms on the right-hand side of this equation can be rearranged to yield

$$P = a + b(A)(M) \tag{8.7}$$

If you compare Equation 8.7 with Equation 8.3, you will note that Equation 8.7 is simply a linear function, so performance is assumed to be a linear function of the product of $(A)(M)$. Not only do the constants a and b take into account the different metrics, but the value of b also provides substantive information as well; namely, it indicates how much change in performance (GPA) one expects given a 1-unit increase in the value of the product term $(A)(M)$.

Figure 8.18 plots the relationship between performance and motivation at three different levels of ability based on Equation 8.7, where values of *a* and *b* have been empirically determined from data collected for a sample of 90 students. In this example, $a = -2.0$ and $b = .0015$. The slope of *P* on *M* for any given value of *A* is *bA*. There are several features of this plot worth noting. First, note that the effect of motivation on performance is more pronounced as ability increases. This is evident in the steeper slope ($bA = .165$) for the two variables when the ability score is 110 as compared with the slope when the ability score is 100 ($bA = .150$), and, in turn, as compared to the slope when the ability score is 90 ($bA = .135$). These differences in slope may seem small but they are probably substantial. For example, when the ability score is 110, a 10-unit change in motivation is predicted to yield a (.165)(10) = 1.65-unit change in GPA; when the ability score is 100, a 10-unit change in motivation is predicted to yield a (.150)(10) = 1.50-unit change in GPA; when the ability score is 90, a 10-unit change in motivation is predicted to yield a (.135)(10) = 1.35-unit change in GPA.

Second, note that at each of the different levels of ability (90, 100, and 110), the relationship between motivation and performance is assumed to be linear. Is this a reasonable assumption? Perhaps not. Perhaps the relationship between performance and motivation at a given ability level is better captured by an exponential function in the form of one of the curves in Figure 8.7a. For example, when motivation is on the low end of the motivation metric, increasing it somewhat may not have much impact on performance—it will still be too low to make a difference on performance. But at higher levels of the motivation metric, increasing it will have an impact on performance. This dynamic is captured by the exponential functional forms illustrated in Figure 8.7a. Or

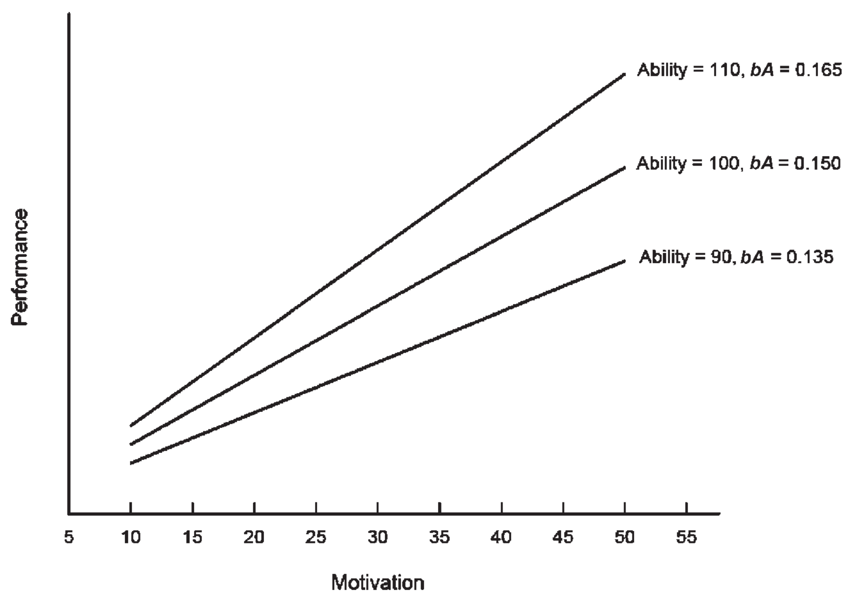


FIGURE 8.18. Example for Performance = Ability × Motivation.

perhaps a power function in the form of one of the curves in Figure 8.9a is applicable. Power functions have the same dynamic as the exponential function in Figure 8.7a, but they "grow" a bit more slowly. Or perhaps an S-shaped function in the form of the curve in Figure 8.17 applies, with floor and ceiling effects on performance occurring at the low and high ends of motivation, respectively.

The multiplicative model specified by Equation 8.6 assumes what is called a *bilinear interaction* between the predictor variables; that is, it assumes that the relationship between the outcome and one of the predictors (in this case, motivation) is always linear no matter what the value is of the other predictor (in this case, ability). To be sure, the value of the slope for the linear relationship between P and M differs depending on the value of A (as noted earlier), but the function form is assumed to be linear. One can modify the model to allow for a nonlinear relationship between performance and motivation at different levels of ability, say, in accord with a power function, as follows

$$P = a + b(A)(M^c) \tag{8.8}$$

where c is an adjustable constant whose value is estimated from data. This model allows for the possibility of a function form like those of Figure 8.9a.

Another notable feature of Figure 8.18 is that at the lowest value of motivation, there is a small degree of separation between the three different lines. The amount of separation between the lines reflects the differences in the effect of ability (at values of 90 vs. 100 vs. 110) on performance when motivation is held constant at the same value. But perhaps the amount of separation should be a bit more or a bit less than what is modeled in Figure 8.18. Equation 8.9 can be further modified to allow for a different amount of separation between the lines than what Equation 8.8 implies, as follows:

$$P = a + b(A)(M^c) + dA \tag{8.9}$$

where d is an adjustable constant whose value is estimated by data. The logic of adding this term is developed in Appendix 8B and is not central to our discussion here. The main points we want to emphasize are the following:

1. The rather simple theoretical representation in Equation 8.6 has nontrivial conceptual ramifications by specifying that the relationship between performance and the predictor variables is captured by the dynamics of a bilinear interaction when, in fact, the interaction may have a different functional form.
2. When building a mathematical model, the metrics of the variables usually have to be addressed (although our next example illustrates a case where this is not necessary).
3. There may be multiple features of the model (e.g., the separation between curves at different levels of one of the component terms as well as the shape of these curves) that must be specified that are not always apparent in simple representations such as Equation 8.6.

The fact is that the often presented model of Performance = Ability × Motivation is poorly specified, and applying principles of mathematical modeling helps to produce a better-specified theory that makes implicit assumptions explicit and highlights complexities that should be taken into account. Appendix 8B develops modeling strategies for this example in more detail and illustrates a substitution principle for building mathematical models. For more discussion of the assumptions of bilinear interactions, see Jaccard and Turrisi (2003).

AN EXAMPLE USING COGNITIVE ALGEBRA

Another example of using mathematical models to represent social phenomena involves models of cognitive algebra. This example illustrates how the implications of a mathematical representation can be pursued without recourse to such things as adjustable constants and complex modeling of data.

Suppose we describe the personal qualities of a political candidate to a person that he or she has not heard of by providing the person with three pieces of information. Suppose that the three pieces of information are all quite positive (e.g., the candidate is said to be honest, smart, and empathic). For purposes of developing this example, suppose we can characterize how positive each piece of information is considered to be using a metric that ranges from 0 to 10, with higher numbers reflecting higher degrees of positivity. We refer to the positivity of a piece of information as P_k, where k indicates the specific piece of information to which we are referring: P_1 refers to the perceived positivity of the first piece of information, P_2 refers to the perceived positivity of the second piece of information, and P_3 refers to the perceived positivity of the third piece of information. Suppose we want to predict how favorable a person will feel toward the candidate based on these three pieces of information. If we let F refer to a person's overall feeling of favorability toward the candidate, with higher values indicating higher levels of favorability, then one model that describes the impact of the information is the following:

$$F = P_1 + P_2 + P_3 \tag{8.10}$$

This model is a simple summative function that specifies that the overall feeling of favorability toward the candidate is the sum of the judged positivity of each individual piece of information (we ignore, for the moment, the metric of F and the issue of adjusting for metric differences). Equation 8.10 can be stated in more general form using summation notation as follows:

$$F = \sum_{i=1}^{k} P_i$$

where k is the number of pieces of information, in this case 3.

Now suppose that instead of a summative function, an averaging function is operating. That is, the overall feeling of favorability is the *average* of the positivity of the information presented rather than the sum of it. In this case, Equation 8.11 becomes

$$F = (P_1 + P_2 + P_3)/3 \tag{8.11}$$

and this can be represented more generally in summation notation as

$$F = \left(\sum_{i=1}^{k} P_i \right) \Big/ k \tag{8.12}$$

What are the implications of specifying the function as being summative versus averaging in form? It turns out, they are considerable. Let's explore the summation model first. Suppose a person judges the positivity values of the three pieces of information as 8, 8, and 8, respectively. The overall feeling of favorability toward the candidate will be 8 + 8 + 8 = 24. Now suppose we describe a second candidate to this person using the same three pieces of information but we add a fourth descriptor to them (cunning), that is judged to have a positivity value of 4. According to the summation model, the overall feeling of favorability toward this new candidate will be 8 + 8 + 8 + 4 = 28, and the person will prefer the second candidate to the first candidate. Psychologically, it is as if the second candidate brings all the same qualities as the first candidate (i.e., P_1, P_2, and P_3) and then "as a bonus," you get a fourth positive attribute as well (P_4). Hence, the person prefers the second candidate to the first candidate.

Now consider instead the averaging function. The overall feeling toward the first candidate is predicted to be (8 + 8 + 8)/3 = 8.0 and the overall feeling toward the second candidate is said to be (8 + 8 + 8 + 4)/4 = 7.0. In the averaging model, exactly the reverse prediction is made in terms of candidate preference; namely, the person now will prefer the first candidate to the second candidate. Psychologically, the first candidate has nothing but very positive qualities, whereas the second candidate has very positive qualities but also some qualities that are only somewhat positive. The person prefers the first candidate, who has nothing but very positive qualities, to the second candidate, who has very positive qualities but also moderately positive qualities.

Which function better accounts for the impressions people form? It turns out that this can be evaluated in a simple experiment in which two candidates would be described, one with three very positive qualities (Candidate A) and a second with three very positive qualities and a fourth moderately positive quality (Candidate B). Participants would then be asked to indicate which of the two candidates they prefer. The summation model predicts that participants should prefer Candidate B to Candidate A, whereas the averaging model predicts that participants should prefer Candidate A to Candidate B. One can differentiate the two models empirically by conducting the above experiment and determining which candidate tends to be preferred. This is a simple experiment without complex modeling. If the results showed that people tend to prefer Candidate A to Candidate B, then this would be consistent with (but not proof of) a summative process rather than an averaging process. If the results showed that people tended to prefer Candidate B to Candidate A, then this would be consistent with (but not proof of) an averaging process rather than a summative process. Which process operates has implications for the design of political campaigns and advertising strategies to sell products. For example, if an advertising campaign adds to a person's cognitions a moderately positive piece of information about a product that is already quite positively evaluated, in the case of the averaging model, the advertisement should backfire and lower evaluations of the target product, thereby adversely affecting sales.

The literature on impression formation has extended these simple models of "cognitive algebra" to more complex model forms. For example, it is almost certainly the case that some information is more important to people in forming impressions than other information. As such, it makes sense to weight each piece of information by its importance to the individual. Equation 8.10 can be modified to include such weights, as follows:

$$F = w_1P_1 + w_2P_2 + w_3P_3 \tag{8.13}$$

where w_i is the importance of information i to the individual. Note that Equation 8.10 is a special case of Equation 8.13, namely the case where $w_1 = w_2 = w_3 = 1$. Expressed in summation notation, Equation 8.13 can be represented as

$$F = \sum_{i=1}^{k} w_iP_i \tag{8.14}$$

For the averaging model, introducing importance weights yields the following:

$$F = (w_1P_1 + w_2P_2 + w_3P_3)/(w_1 + w_2 + w_3) \tag{8.15}$$

Note that Equation 8.11 is a special case of Equation 8.15, namely the case where $w_1 = w_2 = w_3 = 1$. Equation 8.15 can be restated using summation notation as

$$F = \sum_{i=1}^{k} w_iP_i \Big/ \sum_{i=1}^{k} w_i \tag{8.16}$$

By extending the logic of algebraic models to the domain of "cognitive algebra" (which uses the premise that mental operations can be modeled by simple algebra), a great many insights into human information processing have been gained. Some of this research has involved simple experiments that pit competing predictions of different algebraic models against one another, whereas other research has taken the path of more complex math modeling with adjustable constants, error terms, and the like.

Parenthetically, the research literature finds support for both the summation and averaging models. In some contexts, people average the implications of information, whereas in other contexts, they sum it. There also are individual differences in these tendencies, with some people tending to average information in general whereas others tend to sum it in general. There are contexts for which simple summation or averaging models do not hold, and more complex combinatorial models are required to capture the integration dynamics. Interested readers are referred to Anderson (1981) and Fishbein and Ajzen (1975).

AN EXAMPLE USING ATTITUDE CHANGE

As a third example of a mathematical model, we consider a model of attitude change from the communication literature that was developed by Fishbein and Ajzen (1975). The model concerns the case where a source is trying to persuade the recipient of a persuasive message to change his or her belief in something. A belief is conceptualized as a subjective probability that ranges from 0 to 1.00, much like a probability in mathematics. For example, people might believe with a probability of 0.20 that they will contract lung cancer if they smokes cigarettes. Or people might believe with a probability of 0.30 that a particular brand of toothpaste is the best for fighting tooth decay. In the model there are three probabilities that are of interest: (1) the subjective probability that the recipient holds prior to receiving the persuasive message, P_0, (2) the position that the recipient perceives the source takes in his or her persuasive message, also reflected by a subjective probability, P_S, and (3), the subjective probability of the recipient *after* hearing the persuasive message, P_1. For example, the recipient might have an initial belief corresponding to a subjective probability of 0.20, perceive the source as arguing that the target belief should have a subjective probability of 0.70, and after hearing the arguments of the source, the recipient revises his or her subjective probability to be 0.60. These three variables, P_0, P_S, and P_1, are measured variables in the theoretical system.

The amount of belief change that occurs is the difference in subjective probabilities before and after the message, or $P_1 - P_0$. It is the central outcome variable. Fishbein and Ajzen were interested in understanding factors that impact how much belief change occurs, so they constructed a mathematical model to reflect the underlying dynamics. Let *BC* represent belief change and be formally defined as $P_1 - P_0$. Fishbein and Ajzen begin by assuming that the amount of belief change that occurs is a function of the discrepancy between the recipient's initial position and the perceived position of the source—that is, $P_S - P_0$. If a source argues in favor of the exact same position of the recipient, then $P_S - P_0 = 0$, and no belief change will occur. It is only when the source takes a position that is discrepant from the recipient's that belief change can occur. The more discrepant the position taken by the source relative to the recipient's initial position, the greater the potential for belief change. We thus begin with a simple model based on a difference function:

$$BC = (P_S - P_0) \tag{8.17}$$

Not everyone will accept the arguments in a persuasive message. People differ in the likelihood that they will accept a message, with some people having a low probability of message acceptance, others having a moderate probability of message acceptance, and still others having a high probability of message acceptance. Fishbein and Ajzen introduced a parameter into the model to reflect the probability that a recipient would accept the arguments of a message; this parameter is signified by P_A. Equation 8.17 thus becomes

$$BC = P_A(P_S - P_0) \tag{8.18}$$

with the constraint that P_A must range from 0 to 1.0 to reflect a probability metric. If a person completely accepts the message, then $P_A = 1.00$ and the amount of belief change will equal the discrepancy between the recipient's initial position and the position the recipient perceives the source as taking. If a person completely rejects the message, then $P_A = 0.00$ and there is no belief change. If the person is somewhat accepting of the source's message (i.e., P_A is somewhere between 0.00 and 1.00) then the amount of belief change is proportional to P_A.

Next, Fishbein and Ajzen address factors that impact the probability of acceptance of a message. One important factor is how discrepant the message is from the recipient's initial position. In general, people are more likely to accept messages that argue in favor of their existing beliefs as opposed to messages that argue against their existing beliefs. If we let D represent the absolute discrepancy between the recipient's initial position and the perceived position of the source (i.e., $D = |P_S - P_0|$), then the probability of acceptance can be modeled as

$$P_A = (1 - D) \tag{8.19}$$

Note that when $D = 0$, the source is arguing the same position that the recipient already believes and the probability of acceptance is 1.00. As the source's message becomes increasingly discrepant from the recipient's initial position, the probability of acceptance decreases to a minimum of 0.00.

Fishbein and Ajzen recognized that there are factors that can facilitate the acceptance of a message independent of message discrepancy. For example, if the source is a trustworthy and credible person, the exact same message may be more likely to be accepted than if the source is untrustworthy or lacks credibility. Fishbein and Ajzen introduced an adjustable constant to reflect these facilitating conditions, which they labeled f. Equation 8.19 was modified to appear as

$$P_A = (1 - D)^{1/f} \tag{8.20}$$

with the constraint that f be greater than 0. Fishbein and Ajzen thus use a power function to capture the underlying dynamics, where $1/f$ is an adjustable constant. Figure 8.19 presents sample curves for the probability of acceptance as a function of D at different values of f. Note that when $f = 1$, the relationship between the probability of acceptance and message discrepancy is linear with an intercept of 0 and a slope of 1. As f exceeds 1, the probability of acceptance increases rapidly at lower levels of discrepancy and remains high even as message discrepancy increases. As f decreases in value from 1, the

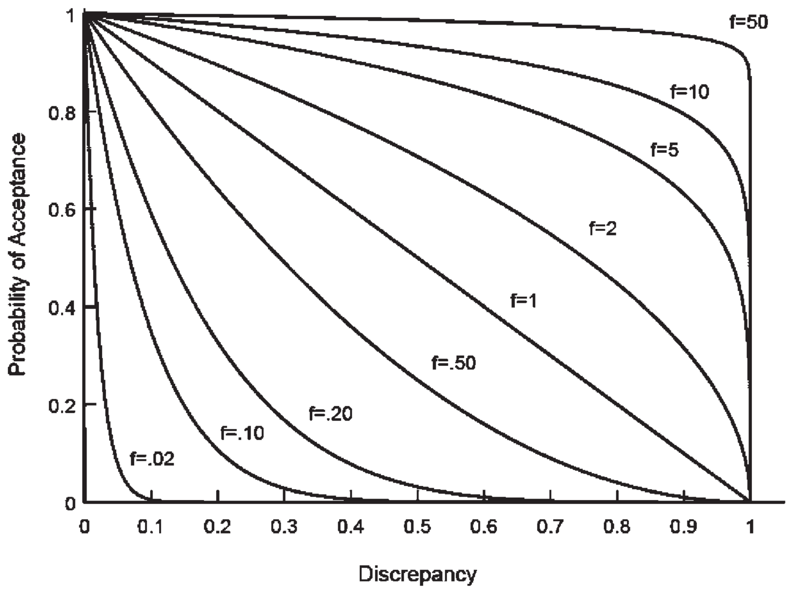


FIGURE 8.19. Fishbein and Ajzen model for probability of acceptance.

probability of message acceptance decreases rapidly at lower levels of discrepancy and remains low as message discrepancy increases.

Equations 8.18 and 8.20 can be combined to yield a single equation. Starting with Equation 8.19, we have

$$BC = P_A(P_S - P_0)$$

Since $PA = (1 - D)^{1/f}$, we can substitute the right-hand side of Equation 8.20 for PA, which yields

$$BC = (1 - D)^{1/f}\,(P_S - P_0)$$

and since $D = |P_S - P_0|$, further substitution yields

$$BC = (1 - (|P_S - P_0|))^{1/f}\,(P_S - P_0)$$

The belief that a person has after hearing a persuasive message can be further specified by recognizing that $BC = (P_1 - P_0)$, so that if we subtract P_0 from both sides of the equation, we obtain

$$P_1 = [(1 - (|P_S - P_0|)^{1/f}(P_S - P_0)] - P_0 \qquad (8.21)$$

Equation 8.21 is a mathematical model that predicts the belief that someone holds after hearing a persuasive message. Although it may appear a bit intimidating to the mathematically uninitiated, it is based on reasonable communication principles and is reasonably precise in the functional forms it posits. The model makes use of observed measures as well as adjustable constants and incorporates a power function. In empirical applications, P_0, P_1, and P_S are measured variables and f is an adjustable constant whose value is estimated from data. The value of f is expected to vary across contexts where factors that facilitate message acceptance vary. For further discussion of this model and its implications, see Fishbein and Ajzen (1975).

AN EXAMPLE USING A TRADITIONAL CAUSAL MODEL

Another example of mathematical modeling in the social sciences is captured by an approach called structural equation modeling (SEM). Although some scientists do not think of structural equation models as mathematical models, they have all the characteristics of mathematical models as described in this chapter. To be sure, they are stochastic rather than deterministic, but their essence is mathematical in nature. The causal model we represent mathematically is presented in the path diagram in Figure 8.20. The model includes disturbance terms (because it is stochastic). We use generic labels for the variables for ease of notation. For this example, we assume that all of the relationships are linear, which is a typical assumption in SEM applications. Each endogenous variable is assumed to be a linear function of all variables that have an arrow

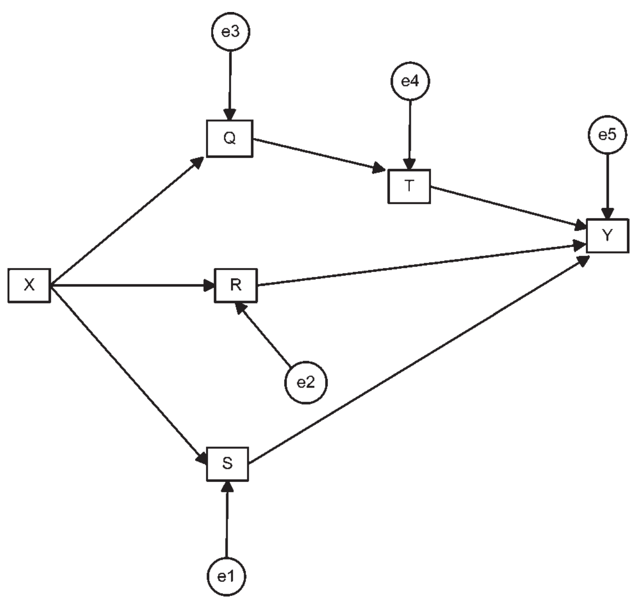


FIGURE 8.20. Causal model.

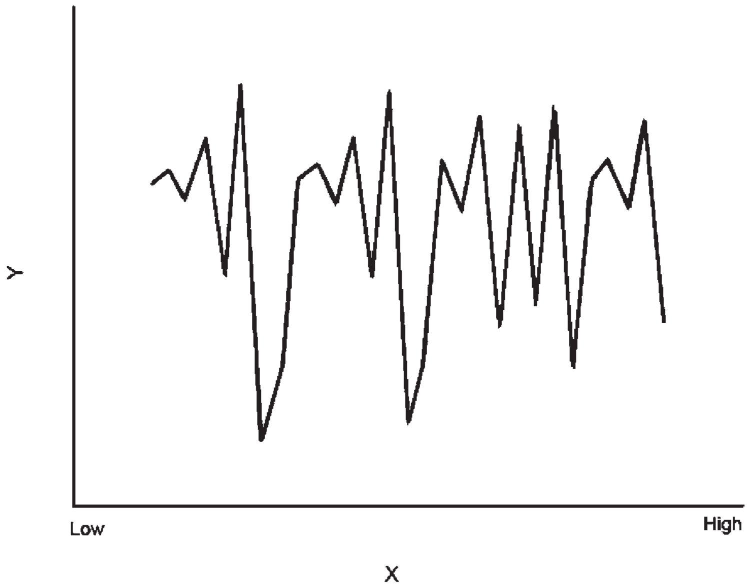


FIGURE 8.21. Chaos theory example.

going directly to it. The model can be expressed as a set of linear equations that are as follows:

$$
\begin{aligned}
Y &= a_1 + b_1T + b_2R + b_3S + e_5 \\
T &= a_2 + b_4Q + e_4 \\
Q &= a_3 + b_5X + e_3 \\
R &= a_4 + b_6X + e_2 \\
S &= a_5 + b_7X + e_1
\end{aligned}
$$

where a_1 through a_5 are adjustable constants representing intercepts, b_1 through b_7 are adjustable constants representing slopes, and e_1 through e_5 are error (or disturbance) terms. The equations yield a model that is overidentified, although constraints must be introduced for estimating the parameters in the presence of the error terms, and other ancillary modeling details must be attended to as well (see Bollen, 1989).

The adjustable constants for the slopes in this model reflect the predicted change in the outcome variable given a 1-unit change in the variable associated with the constant. The one qualification to this statement is for the equation with multiple variables in the linear function. For this equation (where the outcome variable is Y), the slope adjustable constant associated with a given variable in the function is the predicted change in the outcome variable (Y) given a 1-unit change in the variable *holding constant all other variables in the linear function*. In practice, data on each of the variables would be collected and the model would be fit to the data to determine if it could account for the observed data. The data would be used to estimate the values of the adjustable constants so as to maximize model fit. If the fit is reasonable, then values of the adjustable constants are subjected to meaningful interpretation.

This example illustrates another strategy in mathematical modeling when dealing with multiple variables: the modeler creates a systems of equations rather than a single equation to represent the multivariate dynamics.

CHAOS THEORY

An area of mathematical modeling that is receiving increased attention in the social sciences is that of chaos theory. In normal parlance, *chaos* refers to disarray. In the field of chaos theory, this also is true but something systematic is thought to underlie the chaos; what appears chaotic actually has a systematic function generating it. The task of the theorist is to map this function.

Chaos theory is typically applied to changes in systems over time, with the state of a system at time $t + 1$ being impacted by the state of the system at some previous time, t. As an example, consider the simple function

$$X_{t+1} = 1.9 - X_t^2 \tag{8.22}$$

where $t + 1$ refers to the time period following time t. For example, perhaps the time interval in question is a week and suppose that the value of X at time t is 1. Then applying Equation 8.22, the value of X 1 week later (i.e., at time $t + 1$) should be $1.9 - 1^2 = 0.9$. At week 2, this input value is substituted into the right-hand side of Equation 8.22 and the result is the predicted value of X at week number 3. It is $1.9 - .9^2 = 1.09$. To predict the value at week 4, the previous value is again substituted into the right-hand side of Equation 8.22 and the result is $1.9 - 1.09^2 = 0.712$. And so on. The pattern of data is plotted in Figure 8.21, which plots the value of X at each week in a series of weeks. The pattern appears to be unsystematic and chaotic with large swings in values. But note that the underlying process is anything but haphazard. The data were the result of a clearly specified and simple function (Equation 8.22). There was no random error in the system. Rather, the "disarray" was systematically generated. The task of the chaos theorist is to identify patterns that appear to be chaotic and to find the function that generates that "chaos."

In math modeling the term *difference equation* refers to the case where a variable at time t is a function of a variable at time t - 1. If the variable at time t is a function of the immediately preceding point in time, it is called a *first-order difference equation*. If it is predicted from time $t - 2$, it is called a *second-order difference equation*. If it is predicted from time $t - 3$, it is a *third-order difference equation*. And so on.

Chaotic modeling tends to require precise measurement, and results can be dramatically influenced by the slightest "noise" or measurement error in the system. Current analytic methods for chaos models tend to require large numbers of observations. Although chaos theory is typically applied to the analysis of systems across time, the

BOX 8.1. Reading Mathematical Models

When reading mathematical models, social scientists with more limited mathematical backgrounds sometimes feel intimidated by the presence of equations. Because equations make clear and unambiguous statements about the presumed relationships between variables, you should embrace equations, not avoid them. When confronted with an equation that seems complex, here are some things you can do to help work your way through it. First, make a list of the variables in the equation and a list of the adjustable constants. Make sure that each of the variables in your list is clearly defined and that the metrics of the variables are specified. Second, determine if the equation contains any of the major functions we discussed. For example, is a power function present? Is an exponential function present? Is a logistic function present? Once you recognize a familiar function form and you have a sense of the family of curves associated with it, then the substantive implications of the equation should start to become apparent. Remember, the fundamental form of the function can be altered using transformations, so be sensitive to the presence of a function that has a transformation imposed on it. Sometimes the function is "disguised" by the adjustable constants attached to it. Third, for each adjustable constant, think about what it is accomplishing and why it was included in the equation. Is it just a scaling factor, or does it have substantive interpretation, like a slope in a linear relationship? Finally, you can use your favorite statistical package (e.g., SPSS) or graphics software to apply the equation to hypothetical data you generate and then examine the curve graphically and see what happens to it as you change values of the adjustable constants or change the hypothetical data used to generate it in systematic ways. Also, keep in mind the conditional nature that multiplicative functions imply; that is, when you see the multiplication of two variables in an equation, then the size of the derivative (i.e., the size of the effect) of one of the variables in the product term is dependent on the value of the other variable in the product term.

If you encounter mathematical symbols with which you are not familiar, then you can usually find their meaning on the Internet. Below are some commonly encountered symbols. A useful website for learning about many areas of mathematics at many different levels is called "Ask Dr. Math": *mathforum.org/dr.math.*

Common Symbols That Reflect Important Numbers

π = the ratio of the circumference to the diameter of a circle, the number 3.1415926535 . . .

e = the natural logarithm base, the number 2.718281828459 . . .

γ = the Euler–Mascheroni constant, the number 0.577215664901 . . .

ϕ = the golden ratio, the number 1.618033988749 . . .

∞ = infinity

cont.

Symbols for Binary Relations

$=$ means "is the same as"
$\neq$ means "is not equal to"
$<$ means "is less than"
$\leq$ means "is less than or equal to"
$>$ means "is greater than"
$\geq$ means "is greater than or equal to"
$\pm$ means "plus or minus"
$\cong$ means "is congruent to"
$\approx$ means "is approximately equal to"
$\simeq$ means "is similar to"
$\doteq$ means "is nearly equal to"
$\propto$ means "is proportional to"
$\equiv$ means "absolute equality"

Symbols from Mathematical Logic

$\therefore$ means "therefore"
$\because$ means "because"
$\ni$ means "under the condition that"
$\Rightarrow$ means "logically implies that"
$\Leftrightarrow$ means "if and only if"
$\forall$ means "for all"
$\exists$ means "there exists"

Symbols Used in Set Theory

$\subset$ means "this set is a subset of"
$\supset$ means "this set has as a subset"
$\cup$ is the union of two sets and means "take the elements that are in either set"
$\cap$ is the intersection of two sets and means "take the elements that are in either set"
$\varnothing$ refers to the empty set or null set and means "the set without any elements in it"
$\in$ means "is an element of"
$\notin$ means "is not an element of"

Symbols for Operations

n! means "the factorial of"
$\sum$ means "the sum of"
$\prod$ means "the product of"
^ means "to the power of"
$\int$ means "the integral of"

cont.

Additional Notations

Greek letters are used to refer to population parameters, Roman, usually italic, letters are used to refer to sample statistics.

A number raised to .5 or to ½ is the same as the square root of the number. A number raised to the power of –1 is the same as the inverse of the number.

properties of space and distance can be used in place of time. Thus, theorists often distinguish between *temporal chaos* and *spatial chaos*. Temporal chaos models that focus on discrete time intervals (e.g., every 10 years; at 3-, 6-, and 12-month intervals) are called *discrete time models*, and those that use time continuously are called *continuous time models*.

A wide range of phenomena is potentially chaotic in nature, including epidemics, economic changes, the stock market, and the mental state of depression, to name a few. However, it is controversial as to whether a truly chaotic system can be isolated in the real world, in the sense described in this chapter (i.e., with a stable, generating function underlying the chaos).

Variants of chaos theory include, among other things, attempts to identify limits of predictable versus unpredictable patterns of data. For example, air flow over the wing of an airplane might be smooth and predictable when the wing is placed at low angles facing the wind. However, the air flow becomes chaotic and unpredictable at larger angles. One could attempt to determine the largest angle that permits smooth air flow, thereby yielding some understanding of this "chaotic system."

The technical aspects of chaos theory are well beyond the scope of this book. However, the theory represents an interesting application of mathematical modeling that promises to have impact in the social sciences in the future.

CATASTROPHE THEORY

Catastrophe theory is another area of mathematical modeling that is receiving attention in the social sciences. A catastrophic event is one where a large and rapid change in a system output occurs even though the system inputs are smooth and continuous. A simple example that captures the idea of catastrophic events is that of increasing the load on a bridge. One can keep adding weight to a bridge and see how the bridge deforms in response to that weight. The deforming of the bridge proceeds in a relatively uniform manner, showing increasing levels of bending. At some critical point, however, additional weight causes the bridge to collapse completely. Phenomena that might be analyzed using catastrophe theory include the occurrence of a nervous breakdown, drug relapse, divorce, a revolution occurring in a society, a demonstration turning into mob violence, or movement from one developmental stage to another in the context of a stage theory of development.

Catastrophe theory, developed by Rene Thom (1975), postulates seven fundamental mathematical equations to describe discontinuous behavior. Catastrophe theory relates outcome variables to what are called *control variables*, which essentially are explanatory variables. The relationships between the variables are expressed mathematically using nonlinear, dynamic systems that rely on different forms of polynomial functions. The spirit of catastrophe theory can be captured intuitively in a model of aggression in dogs as developed by Zeeman (1976). The behavioral outcome ranges from flight to attack, and the response on this dimension is thought to be a function of two emotions that represent control variables: fear and anger. When fear and anger are at their neutral points, then simple increases in either fear or anger lead to a continuous increase in flight or attack responses, respectively. However, if anger is increased in an already fearful dog, then the potential for a sudden jump from flight to attack can occur. Similarly, if fear is increased in an already angry dog, a sudden jump from attack to flight can occur. The mathematical models developed by Thom and expanded by other mathematicians are designed to model such dynamics. Catastrophe theory represents another area of mathematical modeling that is starting to receive attention from the social sciences.

ADDITIONAL EXAMPLES OF MATHEMATICAL MODELS IN THE SOCIAL SCIENCES

Mathematical models exist in all of the major subdisciplines of the social sciences. Most of the subdisciplines have journals that are devoted exclusively to mathematical modeling (e.g., *Journal of Mathematical Psychology*, *Journal of Mathematical Sociology*, *Journal of Quantitative Anthropology, Marketing Science*). Mathematical models also appear in more mainstream journals, but with less frequency. It is impossible to describe the many areas in which mathematical models have been developed, but in this section, we provide a brief sampling to highlight the diversity of applications.

One area where mathematical models have been prominent is in the analysis of human decision making. This endeavor has involved applications of expected-utility theory, linear regression models, Bayesian probability models, and information theory models, to name a few. The models use mathematics to document both the strengths and limitations of humans as information processors when making decisions. Mathematical models also are prominent in theories of memory, learning, language, bargaining, and signal detectability. Mathematical models have been used extensively in the analysis of social networks involving units such as institutions, communities, elites, friendship systems, kinship systems, and trade networks. Mathematical models of political behavior have explored such issues as voting and fairness. Behavior genetics relies heavily on mathematical decompositions of the effects of unique environmental influences, shared environmental influences, and genetic influences on human behavior in the context of twin studies. Spatial models are used to analyze residential and neighborhood patterning and the effects of this patterning on a wide range of phenomena. Geostatistical techniques explore spatial autocorrelation structures and then use mathematical models to estimate values of variables across regions. Our list could go on, but hopefully, this provides you with a sense of the diverse areas to which mathematical models have been applied.

EMERGENT THEORY CONSTRUCTION AND MATHEMATICAL MODELS

It may seem heretical to use the terms "grounded/emergent theory" and "mathematical modeling" in the same sentence, but there is no reason why some of the concepts developed in this chapter could not be used within emergent theory frameworks. For example, as one thinks about the conceptual relationships that emerge from qualitative data, are these relationships linear or nonlinear in form? If nonlinear, might they be described by logarithmic functions, exponential functions, power functions, polynomial functions, sine functions, or cosine functions? Could some systematic combination of variables underlie what seems to be chaos? Is their anything to be gained by thinking about qualitative data in terms of the logic of multiplicative modeling or cognitive algebra? And so on.

We noted earlier that mathematical modelers usually give short shrift to how the variables they decide to include in their models are chosen. Certainly an emergent theoretical framework might help them select their variables in informed and creative ways.

We think it would be interesting to have a mathematical modeler and a grounded/emergent theorist work as a multidisciplinary team on a common problem, with the instructions to develop an integrated finished product that they both would "sign off" on. Such a collaboration would undoubtedly yield nontraditional perspectives on matters.

SUMMARY AND CONCLUDING COMMENTS

Mathematical modeling is an elegant framework for constructing theories. The emphasis of mathematical modeling is thinking in terms of functions and how to describe relationships between variables in mathematical terms. Functions specify how input variables should be operated upon mathematically to produce outputs. One of the most commonly used functions in the social sciences is the linear function, which has two adjustable constants, a slope and an intercept. The intercept is the output value of the function when the input X equals zero, and the slope is the change in the output given a 1-unit increase in X. Rarely do data conform to a perfect linear function. Model disparities are accommodated through the addition of disturbance or error terms to models.

Errors are assumed to be random and inconsequential for the purposes at hand. Models without errors are deterministic, and models with errors are stochastic.

Mathematical models vary in their number of adjustable constants and the meaning of those constants/parameters. Some parameters reflect rates of change in function output per unit change in function input. These rates are best captured using the concepts of derivatives and differentiation from calculus. Derivatives refer to the concept of instantaneous change, and differentiation refers to mathematical methods for calculating the amount of instantaneous change that occurs. Integrals focus on "areas under the curve" or accumulation, and integration refers to the methods used to calculate integrals.

Mathematical models also differ in their identification status, with some models being underidentified, some being just-identified, and others being overidentified. A just-identified model is one for which there is a unique solution for each estimated parameter. In an underidentified model there are an infinite number of solutions for one or more of the model parameters. In an overidentified model there is a unique solution for the model parameters and there also is more than one feature that can be used to independently estimate a parameter value. Finally, mathematical models vary in the metrics upon which they rely upon for the input variables. The metrics of variables can affect the type of functions used to describe the relationships between variables and how the parameter variables are interpreted.

Although the assumption of linear relationships is ubiquitous in the social sciences, nonlinear relationships could very well be more common. Five major classes of nonlinear functions are logarithmic functions, exponential functions, power functions, polynomial functions, and trigonometric functions. Logarithms are used to model growth or change, where the change is rapid at first and then slows to a gradual and eventually almost nonexistent pace. Logarithmic models reflect rates of increase that are inversely proportional to the output value of the function. Exponential functions are the inverse of log functions, with the two functions mirroring each other's properties. Power functions have a similar shape to exponential and logarithmic functions, but differ at higher values of the input X. Power curves eventually outgrow a logarithmic function and undergrow an exponential function. Polynomial functions are the sum of power functions and can accommodate phenomena with "wiggles and turns." The more bends there are in a curve, the greater the number of polynomial terms that are needed to reflect those bends. Trigonometric functions are used to model cyclical phenomena, with the most common functions being the sine and cosine functions.

Functions can be manipulated through transformations and can be combined to form new functions. For example, the often used logistic function is a combination of a bounded exponential function and an increasing exponential function. Combining and manipulating functions is a key ingredient to building effective mathematical models. A typical theory construction process involves breaking up the overall process into a series of smaller component processes, specifying a function to reflect each component, and then assembling the component functions into a larger whole.

When functions involve more than one input variable, additional levels of flexibility and complexity are introduced, as the input variables are combined additively or multiplicatively. With multiple input variables, the theorist often thinks of the function relating each individual input variable to the output variable and then combines the different variables and their functions, while taking into account the synergistic interaction between the input variables. When choosing functions to use in a model, it is advisable not to overparameterize the model or to add parameters that are not subject to meaningful substantive interpretation.

Mathematical modeling represents a sophisticated way of thinking about relationships between variables. The approach is underutilized in the social sciences, and we believe that theory construction efforts can benefit from thinking about phenomena from this perspective.

SUGGESTED READINGS

Abramowitz, M., & Stegun, I. (1974). *Handbook of mathematical functions, with formulas, graphs, and mathematical tables*. New York: Dover.—A classic on a wide range of functions. A bit dated, but still informative.

Aris, R. (1994). *Mathematical modeling techniques*. New York: Dover.—A description of mathematical modeling techniques, with a bent toward the physical sciences.

Bender, E. (1978). *An introduction to mathematical modeling*. Mineola, NY: Dover.—Describing math modeling through hundreds of examples, this book is primarily oriented to the physical and engineering sciences.

Moony, D., & Swift, R. (1999). *A course in mathematical modeling*. New York: Mathematical Associates of America.—A good introduction to mathematical modeling.

Saunders, P. (1980). *An introduction to catastrophe theory*. Cambridge, UK: Cambridge University Press.—A midlevel book on catastrophe theory.

Williams, G. (1997). *Chaos theory tamed*. Washington, DC: Joseph Henry Press.—An excellent presentation of chaos theory.

KEY TERMS

discrete variable (p. 178)
continuous variable (p. 178)
axioms (p. 179)
theorems (p. 179)
function (p. 179)
function domain (p. 180)
function range (p. 180)
linear function (p. 180)
slope (p. 181)
intercept (p. 184)
adjustable parameter (p. 186)
fixed parameter (p. 186)
estimated parameter (p. 186)
derivative (p. 187)

differentiation (p. 187)
integral (p. 190)
integration (p. 190)
deterministic model (p. 186)
probabilistic model (p. 186)
stochastic model (p. 186)
just-identified model (p. 191)
overidentified model (p. 191)
underidentified model (p. 191)
concavity (p. 193)
proportionality (p. 193)
scaling constant (p. 193)
logarithmic function (p. 193)
exponential function (p. 194)
Napier's constant (p. 194)
power function (p. 197)
polynomial function (p. 198)
trigonometric function (p. 200)
rational function (p. 205)
bounded exponential function (p. 206)
logistic function (p. 206)
vertical shift (p. 205)
horizontal shift (p. 205)
vertical crunch (p. 206)
horizontal crunch (p. 206)
sigmoid function (p. 206)
bilinear interaction (p. 213)
chaos theory (p. 222)
discrete time model (p. 225)
continuous time model (p. 225)
temporal chaos (p. 225)
spatial chaos (p. 225)
catastrophe theory (p. 225)
difference equation (p. 222)
first-order difference equation (p. 222)

EXERCISES

Exercises to Reinforce Concepts

1. What is the difference between an axiom and a theorem?
2. What is a function?
3. How do you interpret the value of a slope and intercept in a linear relationship?
4. How do you calculate a slope in a linear relationship? How do you calculate the intercept?
5. Why would you add an error term to a model? How does this relate to the terms *stochastic* and *deterministic* modeling?
6. What is the difference between a derivative and differentiation?
7. What is the difference between a first and second derivative?

8. What is integration?
9. Why are metrics important to consider when constructing a mathematical model?
10. Briefly describe the major types of nonlinear functions.
11. What are the major types of transformations to functions, and what effects do they have?
12. What criteria are used in choosing a function?
13. Briefly characterize chaos theory.
14. Briefly characterize catastrophe theory.

Exercises to Apply Concepts

1. Find an example of a mathematical model in the literature and write a summary of it. Discuss each of the key parameters in the model and what those parameters represent. Develop the model's conceptual and substantive implications.
2. Develop a mathematical model for a phenomenon of interest to you. Begin by identifying your outcome variable and then variables that you believe are related to it. Specify the functions relating the variables and add relevant constants to the equations, as appropriate. Justify conceptually each function and each constant. Decide if the model should be deterministic or stochastic. Start simple and then build complexity into the model, accordingly.
3. Pick a phenomenon of interest to you and try to apply either chaos theory or catastrophe theory to it. Describe the new theory as completely as you can.

Appendix 8A

SPSS Code for Exploring Distribution Properties

This appendix presents syntax from SPSS that can be used to examine curves produced by different functions.

First, open SPSS with the data field blank. We will use the syntax editor. The first step is to create a variable with a large number of cases, say 100,000. This is accomplished with the following syntax:

```
INPUT PROGRAM.
LOOP #I = 1 TO 100000.
END CASE.
END LOOP.
END FILE.
END INPUT PROGRAM.
COMPUTE X = $CASENUM.
EXECUTE.
```

The last entry in the LOOP command (100000) specifies the number of cases to generate. Numbers are generated in a variable called *X*, and these numbers range from 1 to the number of cases generated. These can be transformed to take on any metric you wish. For example, to have them range from 0 to 1, multiply *X* by .00001. To have them range from –5 to +5, multiply by .00001, subtract 0.5, and then multiply the result by 10. And so on.

Next, we compute the function we are interested in graphing. Suppose it is a log to the base 10. SPSS offers numerous built-in functions, and in this case, we use the syntax

```
COMPUTE XX=LG10(X).
GRAPH
/HISTOGRAM=XX.
```

The last two lines construct a histogram of the data, and the shape of the function will be evident from this. You can add adjustable constants and perform various transformations discussed in the chapter, as desired.

The major function commands available in SPSS are arsin, artan, cos, exp, lg10, ln, sin, and sqrt. These are defined in the help menu in SPSS. One also can work with a wide range of statistical functions, including the logistic function. Note that it is possible to calculate a log to any base from the natural log. The logarithm base *a* of any number is the natural logarithm of the number divided by the natural logarithm of the base. For example, to calculate $\log_2(100)$, evaluate the expression $\ln(100)/\ln(2)$.

There are a host of graphic software programs (for both PC and Mac) designed for scientists that allow them to graph a wide range of functions easily and quickly. These include CoPlot, DPlot, Sigma Plot, and Grapher. We are fond of DPlot. Other statistical software programs that have good graphics packages are S Plus, R, and Statistica.

Appendix 8B

Additional Modeling Issues for the Performance, Motivation, and Ability Example

This appendix describes details for the example modeling the effects of ability and motivation on performance, where the relationship between performance and motivation is nonlinear instead of linear at a given level of ability. We assume the reader is versed in standard statistical methods and psychometric theory. We illustrate the case first where motivation is assumed to impact performance in accord with a power function, with the shape of the power function changing as a function of ability. Then we mention the case where the relationship between performance and motivation is assumed to be S-shaped, with the form of the S varying as a function of ability.

We build the power function model by first positing that performance is a power function of motivation,

$$P = a + bM^c \tag{A.1}$$

where a and b are adjustable constants to accommodate metrics and c is an adjustable constant to isolate the relevant power curve in light of a and b. According to the broader theory, the effect of motivation on performance varies depending on ability (e.g., when ability is low, increases in motivation will have negligible effects on performance, but when ability is moderate to high, increases in motivation will have a more substantial impact on performance). Stated another way, the shape of the power curve will differ depending on the level of ability of students, such that the value of c is some function of A. In addition, it is likely the case that the adjustable constants a and b vary as a function of A. To simplify matters and to develop the underlying logic, we will assume that c is a linear function of A, that a is a linear function of A, and that b is a linear function of A. This yields the equations

$$c = d + fA$$
$$a = g + hA$$
$$b = i + jA$$

where c, d, f, g, h, i, and j are adjustable constants that conform to the respective linear models. Using substitution principles, we can substitute the right-hand side of these equations into A.1, which yields

$$P = (g + hA) + (i + jA)(M)^{(d + fA)}$$

Expanding, we obtain

$$P = (g + hA) + iM^{(d + fA)} + jAM^{(d + fA)}$$

We can rewrite this equation using the more familiar symbols of *a* and *b* for adjustable constants in regression analysis:

$$P = a + b_1A + b_2M^{(b_3 + b_4A)} + b_5AM^{(b_3 + b_4A)}$$

This model can be fit to data and the values of the adjustable constants estimated using nonlinear regression algorithms in SPSS or some other statistical package. The adjustable constants are amenable to interpretation, but we forgo explication of this here. Additional interpretative complications present themselves if the metrics involved are arbitrary, but we do not pursue such matters here either (see Blanton & Jaccard, 2006a, 2006b).

One intuitive way of seeing the implications of the function once the values of the adjustable constants are estimated is to calculate predicted scores that vary M by 1 unit at select values of A. These can be graphed and then subjected to interpretation.

An alternative approach to modeling the data that uses methods that are more familiar to social scientists is to use polynomial regression. In this approach, performance is assumed to be a quadratic function of motivation. Although the full quadratic curve most certainly is not applicable (because it is U-shaped), the part of the curve that forms the right half of the "U" could apply. The model includes adjustable constants to isolate this portion. We begin by writing a model where performance is a quadratic function of motivation

$$P = a_1 + b_1M + b_2M^2 \tag{A.2}$$

and the adjustable constants in this equation (the intercept and the regression coefficients) are modeled as being a linear function of ability (we could use a nonlinear function, but for the sake of pedagogy, we assume a linear function), yielding

$$\begin{aligned} a_1 &= a_2 + b_3A \\ b_1 &= a_3 + b_4A \\ b_2 &= a_4 + b_5A \end{aligned}$$

Using the substitution principle, we substitute the right-hand sides of these equations for their respective terms in Equation A.2, which produces

$$P = (a_2 + b_3A) + (a_3 + b_4A)M + (a_4 + b_5A)M^2$$

Expanding this yields

$$P = a_2 + b_3A + a_3M + b_4AM + a_4M^2 + b_5AM^2$$

Rearranging and relabeling the constants to conform to more traditional notation yields the model

$$P = a + b_1A + b_2M + b_3AM + b_4M^2 + b_5AM^2$$

This model can be fit using standard least squares regression.

To model an S-shaped function, one can stay with polynomial regression but extend the logic to a cubic function. The basic idea is to express performance as a cubic function of motivation

$$P = a_1 + b_1M + b_2M^2 + b_3M^3$$

and then to model the adjustable constants as a function of *A*. Finally, use the substitution method to derive the more complex generating function.

Alternatively, one can use a logistic function to capture the S shape and then model the adjustable constants within it as a function of *A*. This approach requires the use of nonlinear algorithms in estimating the adjustable constants.

9

Simulation as a Theory Development Method

In theory, there is no difference between theory and practice. But in practice, there is.

—A Computer Scientist

To this point we have examined how causal thinking and mathematical modeling provide frameworks for theory construction. Another approach that can help in the theory construction process is the use of simulations, particularly when the phenomenon being studied is a complex and dynamic process. Simulations can be used to generate theory in their own right or they can be used as a complement to causal and mathematical modeling.

Most readers likely are aware of the simulations used for training airline pilots and sharpening their ability to handle unanticipated and infrequently occurring contingencies. These devices are gyroscopically controlled rooms constructed so that their interior looks, feels, and operates like the criterion aircraft cockpit to a pilot during the actual in-flight operation of that aircraft. Another familiar form of simulation is the business games and management simulations used by industrial organizations and business schools to train managers and students. Less well known, however, are the large number of simulations that are being pursued by economists, political scientists, psychologists, and sociologists, among others, to gain perspectives on theories of complex social systems.

For example, in 2008, the economy of the United States dropped precipitously into a recession with the collapse of several large banks, insurance companies, and investment firms. The swings in the stock market and the economic responses of investors confounded traditional economic theory that embraced the notion of equilibrium. Equilibrium theory views markets as the product of a balance of forces that respond primarily to new information about, for example, problems a company is having or changes in the housing supply (Farmer & Geanakoplos, 2002). However, the volatility of the markets made clear that something more was at work. Many economists felt that the volatility reflected, in part, the complex dynamics operating between thousands of traders who compete, interact, and trade with one another on a daily basis. Thus, it is not enough to apply the principle of equilibrium. One also must take into account these thousands of relationships to better appreciate market volatility. Economists developed simulations where thousands of artificially intelligent "agents," represented on a computer, were programmed to have different investing orientations, habits, and social and business networks with other traders. The simulations set the agents about their way to interact and transact their business and observations on a daily basis over an extended period of time, the equivalent of, say, several weeks. Observations were then made about how the simulated market behaved across this more extended time period, especially as the researcher introduced new information into the system. The results of such agent-based simulations provided new perspectives on traditional economic theories of equilibrium as well as the effects of other policy-based innovations on aspects of market volatility (Farmer & Geanakoplos, 2002; Westerhoff, 2008). The level of complexity of these simulations, with the hundreds of thousands of interactions between agents that are captured over extended time periods, could not possibly be addressed by traditional methods. By using such agent-based simulations, economists were able to evaluate the effects of changes introduced into the system or the effects of changing one or more underlying assumption about, for example, agent orientations or about agent networks, on higher-level system outputs. This was then used to generate new theories of market volatility.

In this chapter we consider how simulations can be used to develop theory. We begin by defining simulation strategies and then discuss some of the uses of research-based simulations. After describing different types of simulations, we consider the core activity of analyzing criterion systems for purposes of designing simulations. It is at this stage where new ideas and theory clarification often occur. We then discuss the strategies of virtual experiments and agent-based modeling as theory construction strategies and conclude by identifying resources that will facilitate the conduct of simulations.

DEFINING SIMULATIONS

According to *Webster's* dictionary, a simulation is defined as "the imitative representation of the functioning of one system or process by means of the functioning of another." The original system that one seeks to simulate is called the *criterion system*, and the imitative representation of that system is called the *simulation*. The criterion system is generally more complex, whereas the simulation is a simpler representation that incorporates only selected characteristics of the criterion system. Different simulations of the same criterion system may thus assume different layers of complexity, depending upon how many and just which features are incorporated. The criterion systems may be virtually anything, including physical entities, social entities, corporate or organizational entities, economic systems, mental processes, and military maneuvers. Moreover, the simulation system can attempt to represent the characteristics of the criterion system using a variety of forms, including physical, pictorial, verbal, mathematical, and logical approaches. Our attention is confined to simulations developed for studying some aspect of human systems at either the individual or group level.

A distinction can be made between *pedagogical* and *research simulations*. The former are developed primarily to be used as training aids. In contrast, the principal function of research simulations is as a strategy for developing, testing, and extending conceptualizations. Because pedagogical simulations often can be used as research tools, the difference between pedagogical and research simulations generally is one of emphasis. Suppose engineers in the process of designing a new aircraft wanted to determine which of several different instrument panel configurations would generate optimal speed and accuracy of a pilot's response. By installing these configurations in a simulator and using an appropriate sample of test pilots, they could conduct research to answer this question.

THE USES OF RESEARCH SIMULATIONS

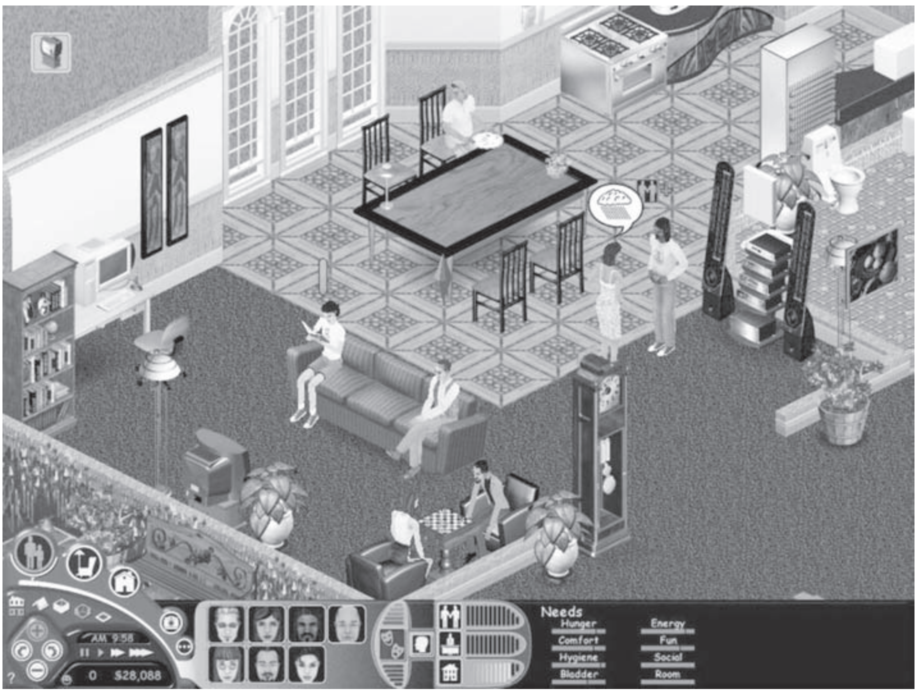
Research-related simulations serve two basic functions: (1) to build and clarify theories, and (2) to test theories. Our emphasis in this chapter is on the former. Developing simulations generally requires the theorist to think about the criterion system in comprehensive terms. Doing so can force confrontation with the inadequacies of one's original conceptualization, thereby revealing the need for more detailed explication of the conceptualization. This, in turn, often results in the expansion or clarification of the original theory. Thus, apart from the data that may be generated when conducting the simulation at some later point in time, the sheer act of developing a research simulation can be viewed as an aid to theory generation and clarification.

THE DIFFERENCE BETWEEN SIMULATIONS AND LABORATORY EXPERIMENTS

Although simulation and experimentation can be integrated in a single study, when not integrated, the characteristics of the typical simulation contrast sharply with those of experimentation. Whereas fully experimental designs involve "tight" investigator control over both the presentation of the stimuli and the sequence of events, simulation allows a tendency for events, their sequence (the "information flow"), and perhaps even the consequences to be determined by the research "participants." Whereas experimentation usually reflects an effort either to eliminate nonfocal factors and/or hold them constant, simulations generally permit these factors to vary freely. Whereas experimentation generally concentrates on a limited number of independent and dependent variables, simulations generally include a greater number of variables, particularly potential causes of the variable of interest. Whereas to achieve their goals, experiments "tie" and "untie" variables (and levels of variables) in ways that may divorce them from everyday reality,

BOX 9.1. Simulations Go Public:
The Sims and Virtual Environments

One of the most popular computer games of all time is *The Sims*, a life simulation computer exercise. The program, developed by Will Wright, is a simulation of the day-to-day activities of avatars, generically called the "Sims" as a play on the word *simulation*. The user creates a virtual neighborhood, with different families inhabiting different houses, and then observes the avatars as they interact and go about their daily duties. The user controls the personality of the individual avatars, the composition of the families, the daily activities of each individual, the consumable goods that are in each household, and a variety of other features of the virtual environment. There is a degree of unpredictability built into the program, as events happen outside the control of the user. The game has no goal or purpose—you do not win. Rather, you simply specify the nature of the environment and the people within it and then watch life play itself out. The user structures the daily activities of the avatars to help them achieve their personal goals. The user makes decisions about how much time the avatars engage in exercise, reading, personal hygiene, eating, what they do for fun, how they deal with depression, their financial dealings, and so on. In newer versions of the game, children age and grow up.



cont.

The user can direct an avatar to interact with another avatar. The Sims speak a language called *Simlish*. Informal Internet groups have formed to analyze and decipher the linguistic structure of Simlish. The game has spawned a large number of expansion packs, including Livin' Large (which adds more events, Sims, careers, and the ability to establish multiple neighborhoods), House Party (which adds party-related content), Hot Date (which allows Sims to meet or pick up other Sims for romantic encounters in a city environment), Vacation (which allows the player to take Sims to various vacation destinations), Unleashed (which gives Sims the ability to train pets and grow crops), and Superstar (which allows Sims to visit a Hollywood-like town).

The screen shot from The Sims displays a family interacting in their house. A diamond is over the head of the avatar sitting with crossed legs on the couch; the diamond indicates that he is the character of focus. At the bottom of the screen is a summary of statistics about the focal character, summarizing his need states.

The Sims is in over 50 million households worldwide as of 2008. In many ways, it captures and illustrates some of the issues with which social scientists will grapple as they create virtual environments and virtual interactions between avatars and humans, and it illustrates the basic idea of programming and then watching the programmed principles play themselves out over time.

simulations strive to keep these factors "tied" in a manner that is consonant with the way in which they are associated in the everyday world criterion system. The presence of more background variables coupled with more complex representations of independent variables produces a "richer" environment out of which sometimes emerge unexpected, or "serendipitous" findings that serve to extend the researcher's conceptualization. Whereas experiments are admirably suited for evaluating the input(s) and outcome(s) of a process, simulations tend to place greater emphasis on the process itself.

Many research simulations are developed and used because they are amenable to manipulations that would be impossible, impractical, or too costly to perform with the criterion system itself. In this sense, they are a form of experimentation and comparable to a laboratory experiment. Simulations are particularly useful when the phenomenon of interest is enmeshed in a complex, dynamic system, and the researcher's intent is to extrapolate from the findings to what is probable under real-world conditions.

BASIC SIMULATION VARIETIES

Simulations can be described in terms of a number of features, and any specific simulation can be identified as possessing some features and not others.

All-Machine versus Person–Machine Simulations

One important distinction is between all-machine and person–machine simulations. *Person–machine simulations* involve one or more human participants as an integral part of the operation of the simulation. In contrast, *all-machine simulations* simulate human behavior and/or mental functioning in the absence of any direct interaction with a human participant. There are two major subvarieties: *physical analogue simulations*, which use some sort of physical model to simulate human dynamics, and *mathematical simulations*, which rely entirely on mathematics and computer programs. Artificial intelligence exemplifies a behavioral science domain where the emphasis is on developing computer programs that resemble some form of intelligent thought.

Descriptive versus Analytic Simulations

A distinction is sometimes made between descriptive and analytic simulations. The former emphasize structure, the latter emphasize process. The object of a *descriptive simulation* is to reproduce the physical structure of the criterion system in as much detail as possible, usually on a smaller scale. In contrast, an *analytic simulation* seeks to reproduce the process of the criterion system. Whether the simulation reproduces the physical appearance of the criterion system is of little importance. "The important factor is that the components and variables being investigated respond in a manner comparable to that of the behavior of the real system" (Dawson, 1963, p. 223). Most research simulations are analytic—that is, the criterion system is dynamic, containing characteristics that operate in sequence and over time—and the simulation attempts to represent this process.

Real-Time versus Compressed-Time versus Expanded-Time Simulations

The focus on process raises another distinction. In *real-time simulations*, the process under study takes the same amount of time as it takes in the criterion system. *Compressed-time simulations* focus on phenomena that are extended in time in the criterion system (e.g., interaction patterns between groups over a 2-year period) and enable these patterns to be studied in a much shorter time period, say, 15 minutes. In contrast, *expanded-time simulations* slow down rapid processes so that they can be more carefully studied.

Deterministic versus Nondeterministic Simulations

Another common distinction is between deterministic and nondeterministic simulations. In *deterministic simulations* the precise outcome is determined by the initial input, whereas in *nondeterministic simulations*, the outcome is determined by some combination of the initial input with the "chance" events that occur during operation of the simulation. In other words, deterministic simulations do not build "noise" into the system, whereas nondeterministic simulations do.

Free versus Experimental Simulations

A related distinction is between free and experimental simulations. The defining characteristic of a *free simulation* is that events that occur during the simulation are shaped, in part or entirely, by the behavior of the participants themselves. In contrast, all information that reaches the participant in an *experimental simulation* is preprogrammed by the researcher. Although participants in the experimental simulation believe that they control their own fate (at least to some degree), in point of fact, they do not. This defining characteristic renders the experimental simulation more like the standard laboratory experiment than like the free simulation because the independent variables remain under complete control of the experimenter (Fromkin & Streufert, 1976, p. 425).

Macro- versus Microsimulations

Other ways of describing simulations involve distinguishing them in terms of their scope. *Macrosimulations* involve large-scale criterion systems that include many people and/or variables. An example would be an attempt to simulate the gold-trading system of the free world. *Microsimulations* involve few people and/or variables. At one extreme, a microsimulation would refer to a single type of process (e.g., acquiring information from the outside world as a basis for reaching a decision) going on in the mind of a single individual.

Content-Oriented Simulation

Simulations are often categorized in terms of the specific domain of content they address. Some of the more popular content-oriented simulations in the behavioral sciences include international simulations (Guetzkow, Akger, Brody, Noel, & Snyder, 1963; Silverman & Bryden, 2007), economic system simulations (e.g., Schubert, 1960; Westerhoff, 2008), small group/social process simulations (e.g., McGrath & Altman, 1966; Suleiman, Troitzsch, & Gilbert, 2000), interpersonal bargaining–negotiation simulations (Deutsch & Krauss, 1960; Luce & Raiffa, 1957), organizational simulations (Lomi & Larsen, 2001), and cognitive process simulations (Jacoby, Jaccard, Kuss, Troutman, & Mazursky, 1987; Newell & Simon, 1972; Sun, 2006), to name a few.

THE ANALYSIS OF CRITERION SYSTEMS AS A BASIS FOR THEORY CONSTRUCTION

When developing a research simulation, one begins by developing a conceptualization (essentially, a theory) as to what constitute the essential features of the criterion system that need to be represented in the simulation. Next, one strives to build a simulation that corresponds to the essential features of that conceptualization. Decisions must be made about which features of the criterion system to include and which features to omit. Once designed, the simulation can be used to test a variety of theoretical propositions and to provide feedback on them. In practice, the first two activities go through several iterations, being refined and extended at each iteration. In part, this is because certain features of the criterion system may be so pervasive that, like water to the deep sea dweller, their importance initially goes unrecognized, becoming apparent only after one tries to implement the initial version of the simulation and realizes that something important is missing. This leads to a refined conceptualization of the criterion system, which, in turn, leads to incorporating additional layers of sophistication into the simulation. Throughout the process of simulation development, the goal is to approximate key features of the criterion system as closely as possible and in as much detail as is necessary to answer the question at hand. It is during this phase of simulation design that theory construction and development often occurs. As the theorist carefully analyzes the criterion system, trying to capture and think through the key elements of it that must be captured in the simulation and those elements that can be omitted, important variables and possible relationships between those variables often become evident. We illustrate two case studies of this phenomenon.

Simulation of Information Accessing in Consumer Purchase Decisions

Jacoby has developed an informal theory of information accessing during decision making and a program of research based on that theory using simulation technology. The initial simulations were developed in reaction to limitations perceived to be inherent in studies Jacoby had conducted on the subject of information overload in consumer decision making. According to the overload hypothesis, by increasing the amount of information to which a decision maker must attend during a brief period of time, a point will be reached beyond which the decision maker's ability to cope with this information is strained and the quality of his or her decision making deteriorates. Conducted in 1971 through 1973 (e.g., Jacoby, Speller, & Berning, 1974; Jacoby, Speller, & Kohn, 1974), the overload studies came about in response to arguments from consumer advocates who maintained that, when it came to communicating with consumers, "more information is better." The overload studies were designed to assess whether this "more is better" assumption, or the competing overload hypothesis, would be empirically confirmed.

The overload studies employed variations of a common methodological protocol. First, questions were used to ascertain each research participant's preferences for various features of a test product. Assume that the test product was ready-to-eat breakfast cereal. Each participant would indicate his or her preferences on such characteristics as type of grain (oat, rice, wheat, etc.), whether it was sweetened or not, whether it contained dry fruit or not, the size (in ounces) of the package, the price, and so on. Next, participants were randomly assigned to the cells of a fully experimental factorial design where one factor was "number of brands" and the second was "number of features described for each brand." Each factor had several levels (e.g., 4, 8, 12, or 16 brands for the first factor, and 4, 8, 12, or 16 features for the second factor). Thus, depending upon the cell to which he or she was assigned, the participant would receive as little as 16 items of information (4 brands × 4 features per brand), or as much as 256 items of information (16 brands × 16 features per brand) to consider. Third, the research participants were instructed to examine and evaluate *all* the information provided to them and use this information as the basis for selecting the brand they would most prefer.

The dozen or so studies that Jacoby conducted found that, regardless of the products tested and whether the participants were undergraduates or adult heads of households, lived in the United States, Puerto Rico, or Germany, the pattern of findings generally reflected an overload effect. That is, the ability to select the brand that most closely matched one's ideal brand improved with modest amounts of information, then decreased as the amount of information increased beyond some optimal level. Reports of these findings began surfacing in most consumer behavior textbooks, in marketing and advertising executive deliberations, and in regulatory agency policymaking. For example, the 1981 Federal Trade Commission Staff Report proposing precautionary language in cigarette labeling cited the overload studies as the basis for its recommendation to have no more than three specific health hazards indicated at a time on cigarette packages, and to rotate these periodically.

Analysis of the Criterion System

As early as 1974, Jacoby began to express reservations about the applicability of these findings to the everyday world of consumer decision making (see Jacoby, 1975, 1977, 1984). The disparity between the conditions of assessment and key features of *in vivo* consumer decision making convinced Jacoby that demonstrating that overload was *possible* using laboratory experimentation provided little basis for contending that it was *probable* in everyday life (the criterion system). To examine whether overload occurred in the field would require developing an empirical system or simulated environment that more closely paralleled and incorporated many of the key features of the criterion system. The objective was to devise a simulated environment that would enable the researcher to examine, how and how much information consumers acquired, when left to their own devices, prior to making a purchase decision. This led to an intensive analysis of the criterion system.

Attention initially was limited to consumer choice behavior as it might occur in the context of a typical full-service supermarket, with the combination of consumer plus environment constituting the criterion system. This system was conceptualized as consisting of (1) a decision maker who, at least in part, relies on information acquired from the outside world; (2) a task environment containing a set of choice options (e.g., different brands of breakfast cereal), each of which has associated information (e.g., price, size, ingredients); and (3) an objective that the decision maker brings to the choice context (e.g., "Out of those available, select the breakfast cereal I would most like to buy"). Considering how these three components typically played out *in vivo*, the following features emerged as key characteristics of the criterion system.

First, information exposure in the criterion system was *consumer-determined*, not experimenter-determined. Consumers in the real world are *active*, not passive, information seekers. That is, they do not enter supermarkets and passively wait to be force-fed certain fixed quantities of information (as was characteristic of the overload experiments). Rather, they actively seek and attend to information on their own. Hence, to better understand what was probable in the real world, the simulation had to permit the participant an active stance in acquiring information from an available pool of information.

Second, the task environment is saturated with information that can be used during decision making. Consider the following. The consumer usually has a choice of supermarkets and grocery stores, with each of these outlets containing an overlapping but somewhat different assortment of products. Full-service American supermarkets of that era typically contained more than 25,000–35,000 different items on their shelves, almost all of which appeared in packages that contain considerable amounts of information. For example, the typical American supermarket contains at least 50–80 different brands of breakfast cereal (out of more than 150 such nationally available brands). Even if one disregards all the photographs and drawings, most breakfast cereal boxes contain more than 100 separate items of information. The *information-intensive* nature of this environment is rendered even more complex by the fact that, in many product categories, the available brands (e.g., Crest toothpaste, Coca-Cola) come in different sizes, flavors, colors, etc.

Third, to avoid being overwhelmed by this complexity, the consumer's prechoice information-seeking behavior reflects great *selectivity*. Consumers generally confine their shopping to one store and, when there, may not go down every aisle. Within any given aisle, they generally do not consider products in every product category. Even within a product category they do consider, as the number of options (i.e., brands, sizes, variations) increases, the consumer is less likely to consider all the available options. And when the consumer does pick up a package for one brand, he or she is not likely to attend to all the information that appears thereon. Thus, consideration of the criterion system suggested that consumers operating *in vivo* are highly selective in regard to how much and just which information they attend to.

Fourth, *memory* plays a critical role in decision making. Consumers need not and likely would not expend time or effort to acquire information from the external environment if they already had that information in mind. When it comes to frequently purchased nondurables, knowing a product's brand name enables one to bring forth from memory a considerable amount of directly relevant information. For example, seeing the brand "Budweiser" might stimulate the consumer to recall such things as the product being a domestic, not foreign, brew, the manufacturer's name (Anheuser Busch), the manufacturer's principal place of business (St. Louis), the packaging (which contains red and white colors and an "A and eagle" logo), the brewing process involving beechwood aging, and around the end of the year, its advertising including Clydesdale horses. Sometimes what is brought forth from memory provides a sufficient basis for arriving at a purchase decision. At other times, we need to update or augment information in memory (e.g., what is the item's current price?) or, when considering something new, acquire information from the external environment as a basis for arriving at our decisions.

Additional reflection suggested a fifth important aspect of the criterion system. All participants in the information overload studies were given information in a set sequence. More than likely, this did not correspond to the sequence in which participants would have acquired the same information had they been left to their own devices. Since the scholarly literature contained a sufficient number of reports to show that the *sequence* in which information is obtained could be important, it was considered necessary that the simulation permit participants to assess information in any sequence they desired.

Thus, as the basis for developing the initial simulation, the criterion system (consisting of both the external information environment and the consumer's internal information processing) was conceptualized as reflecting the following key characteristics: In arriving at his or her purchase decision, the consumer relies upon both internal and external sources of information and, especially in regard to the latter, is active and selective in determining just which, how much, and in what order information is acquired. This analysis also led to the recognition that, given the vast amounts of information available in the decision environment (say, a supermarket or drug store), one could not study the entire environment but needed to focus instead on a limited portion of that environment (i.e., a product category, e.g., toothpaste). Attention thus was confined to studying whether the information overload effect would surface—as it had in the laboratory experiments—when research subjects were permitted to operate in ways that more closely paralleled their behavior in selecting a brand of toothpaste or shampoo in naturalistic settings.

An Early Simulation

Jacoby developed these simulations before personal computers were available on a widespread basis, so several rather crude versions of the simulated environment were implemented relative to the simulations now in place. One early simulation consisted of a "strip board" device that presented participants with a matrix of information about different brands (e.g., of toothpaste or of cereal). The brands were represented in columns (e.g., the first column was labeled Brand A, the second column was labeled Brand B, and so on), and the rows were represented as information dimensions (e.g., price, size). The cells of the matrix contained the specific information about a given brand and a given information dimension where the row and column intersected. Although the row and column labels for the brands and information dimensions were visible, the information in the cells of the matrix was covered by horizontal strips of opaque tape. Each strip ran horizontally, covering the brand-specific information for all brands on a single type of information. For example, one strip covered the prices for each of the available brands;

another strip covered the size, and so on. Research participants could remove a strip, one at a time, thereby revealing the information in the cells for all brands on that type of information; for example, pulling off the tape strip over the price information would reveal the prices in the cells for all the brands. The simulation participants were told that they could acquire as much or as little of the information as desired, removing none, one, several, or all of the tape strips before making their choice about the brand they would choose. Further, they were told they could do so in any order they desired. In some simulations, information identifying the brand name (e.g., Crest) was available in the information pool (so that, if acquired, this information would enable the participants to access information about the brand and manufacturer from their memory). In other simulations, this type of information was absent from the pool.

Although crude, the strip board simulations represented significant improvements relative to prior research strategies. First, although scholars generally agreed that decision making was a *dynamic process* that unfolded over time, almost all of the research on decision making conducted to that point relied on static, cross-sectional, or simple pre–post assessment methodologies. The prior overload research was typical of this genre. In a sense, it was "outcome research," not "process research." Crude as it was, the strip board device enabled information acquisition to be studied dynamically, as it occurred, rather than relying on cross-sectional verbal assessments made before (e.g., "What information do you intend to use?") or after the fact (e.g., "What information did you use?").

Second, comparing poststudy debriefings with the information that the simulation participants acquired during the simulation led to greater appreciation of the fact that not all aspects of decision making operate at consciously retrievable levels. As information processors, we may attend to information of which we had no prior knowledge and therefore could have had no intention to seek out or attend to. Moreover, evidence suggests that we are not capable of recalling all the information which we do acquire, or the sequence in which this information is acquired (see McGuire, 1976; Nisbett & Wilson, 1977). Without having to rely on fallible memory or verbal reports about search behavior, the simulation enabled us to directly assess information-accessing *behavior* as it unfolded. It also enabled us to directly examine just which information was ignored instead of having to ask potentially leading questions to make such determinations.

It is interesting to note that these features of the simulations were viewed as a strength by some methodologists and a weakness by others. On the one hand, the attempt to simulate core features of the criterion system led to a more accurate representation of the role of the information acquisition process in the real world. On the other hand, many facets of information acquisition were out of the experimenter's control (e.g., the amount of information accessed, the sequence in which information was accessed), making it more challenging to make causal inferences from the collected data.

Iterating Simulations and Iterating Theories

As noted earlier, when building a simulation, it often is necessary for a researcher to go through several trial-and-error steps, what we call *iterations*, in order to obtain a reasonably faithful representation of the criterion system. In the process, new variables to include in the system might suggest themselves and new relationships to focus on also might become salient. As one iterates the system, one also gains theoretical insights, and revisions to the theory often occur as well. In this way, not only does the simulation become an improved representation of the criterion system, but theory becomes "iterated" and improved as well. Although theoretical revisions based on data are a natural part of all methods of science, a unique facet of this dynamic for simulation research is that the scientist is not only concerned with improving theory but also with improving the simulation per se, so it can be used that much more effectively in future research.

To illustrate this iterative process, consider the strip board simulation and limitations that were revealed by implementing it. One limitation that became apparent was with the restrictions it placed on information acquisition. Although the strip board enabled us to study the acquisition of types or categories of information (e.g., flavor, price), it did not permit us to analyze the acquisition of single pieces of information associated with a specific brand (e.g., the flavor of Crest toothpaste). Subsequent developments, especially computerized versions of the simulation, circumvented this problem. Research with the strip board device also failed to accommodate the fact that real-world decision makers generally live with the consequences of their decisions. This led to a further change in the simulation to include authentic consequences. For example, simulation participants were told in advance that they would receive several "cents-off" coupons that could be redeemed at local stores, but only for the brand they selected. The simulation thus incorporated an element of *consequentiality*, an important feature of the criterion system.

Spurred partly by conceptual considerations as well as the availability of more powerful computers, the simulation continued to undergo several theory–simulation–theory iterations. For example, unlike purchase decisions made in supermarket settings—which involve consumers attending primarily to two dimensions of the external information environment, namely, "options" (brands) and "properties" (information provided on packaging for brands)—*in vivo* decision making often has decision makers obtaining option and property information from different sources. Hence, the simulation was extended to permit testing of such "properties × options × sources" information environments (e.g., Chestnut & Jacoby, 1980; Hoyer & Jacoby, 1983). Predecision information acquisition rarely occurs in a vacuum. Generally, decision making takes place in the context of important antecedents, such as things the decision maker sees, reads, or hears prior to engaging in decision making. In the consumer context, such antecedents include advertising, pertinent news articles, comments by salespersons, and word-of-mouth communications from family and friends.

The simulation also was extended to examine the impact of such incoming communications on information search (e.g., see Sheluga & Jacoby, 1978). In a modification that may hold the greatest potential both for extending and testing theory, the simulation was further revised to allow scientists to study the *incremental impact of item-by-item information acquisition* on such phenomena as uncertainty reduction (Jacoby et al., 1994), brand evaluations (Johar, Jedidi, & Jacoby, 1997), and attitude formation

(Jacoby, Morrin, Jaccard, Gurhan, & Maheswaran, 2002). Finally, as the simulation was extended to examine information accessing in other types of decision making (e.g., real-world security analysts reaching buy–sell decisions for which there was a hard criterion for evaluating performance; e.g., Jacoby et al., 2002; Morrin et al., 2002), it was found that the theory that had evolved in regard to supermarket decision making had to be revised in several substantial respects, again reflecting the iterative relationship between simulations and theory development.

The above discussion demonstrates how, in developing simulations to test theories, the process of simulation development can contribute to theory development. Building a reasonably faithful simulation requires that one undertake a thorough analysis of the criterion system. Since the simulation is essentially a reflection of one's theory of the criterion system, this analysis often will reveal key assumptions and other facets that help refine and extend the seminal conceptualization. This was the case for Jacoby as he developed the simulation for analyzing consumer information accessing. Thus, even when the motivation for designing the simulation may be the development of a research tool for testing theories, doing so can lead to theoretical insights before testing ever takes place. And once the simulation is tested, recognition of its limitations may also suggest ways for extending one's theory.

Virtual Environments and Avatars

Another example of the use of simulations for theory development is recent work with virtual environments. In this research, participants either wear a head-mounted display that immerses them in a three-dimensional, computer-generated virtual environment, or they view a computer screen that has an engrossing and realistic virtual environment. In some studies, the participant is hardwired so that his or her body movements are shown on the computer screen. The virtual world typically is viewed from a first-person perspective, so that the screen mimics what one would see from one's own eyes. Virtual humans, called *avatars*, can be introduced into the environment, and the researcher controls in real time how the avatars act and react in interactions with the research participant.

The use of avatars and virtual environments is occurring with greater frequency in research in the social sciences, as evidenced, for example, by the creation of a new journal, *CyberPsychology and Behavior.* This research has explored the use of avatars and virtual environments for the study of such diverse phenomena as online counseling, fear of public speaking, dating and interpersonal behavior, treatment of acrophobia (fear of heights), acquisition of medical knowledge by medical residents, consumer purchase decisions, interpretation of facial expressions, nicotine craving, eating disorders, and friendship formation, to name a few. The studies have led researchers to posit basic questions about the generation of virtual environment simulations, such as the amount of visual fidelity that is required for an avatar to seem human, how the avatar should interact with the environment so as to appear natural, the perceptual and memory skills that an avatar should be programmed to display in order to be perceived as normal, and the amount of unpredictability and emotion that should be shown by the avatar so as to appear normal. In essence, a new set of questions about human behavior and perception has emerged in response to the simulation problem of designing a realistic avatar. As social scientists have addressed these questions, they have embarked on collaborations with artists, computer programmers, and graphic designers to form multidisciplinary teams for analyzing the subtleties of human actions and social perceptions—teams that, prior to simulations with virtual environments, simply did not interact with one another. The result has been promising new perspectives on basic mechanisms of perception and human interaction, independent of research that is formally conducted using avatars.

There is no question that with the increasing speed, processing capacity, and power of new computers, virtual environments and avatars will be a useful tool not only in society more generally but in social science research in particular. As simulations become increasingly popular, ways of thinking about the human criterion system will expand accordingly.

SIMULATIONS AND VIRTUAL EXPERIMENTS

Another way in which simulations are used during the initial stages of theory construction is to use all-machine simulations to conduct virtual experiments that provide feedback on initial theoretical propositions. Once the theory is revised and refined so that it behaves well in the virtual experiment, it is applied to real-world dynamics (Davis, Bingham, & Eisenhardt, 2007; Sastry, 1997; Sterman, 2000). As an example, Davis, Eisenhardt, and Bingham (2005) unpacked the global construct of "market dynamism" into four subcategories: velocity, complexity, ambiguity, and unpredictability. They then designed a computer simulation of market systems to examine the effect of each component on market behavior and ran the simulation with three of the components held constant and the fourth construct varying. They did this for each component, in turn, and then amended their initial theory based on the results of the simulation. This theory was then applied to real-world scenarios of market dynamics (see also Repenning, 2002).

AGENT-BASED MODELING

Smith and Conrey (2007) describe a simulation-based approach to theory construction called *agent-based modeling.* This approach uses the spirit of virtual experimentation and was illustrated in the example at the beginning of the chapter on the economic recession in the United States. Agent-based modeling focuses on "agents," who usually are simulated individuals, and variables that describe those agents and their environment. Rather than specifying the causal relationships between agent and environmental variables in the simulation, the emphasis of agent-based modeling is instead on specifying the interactive processes that operate between agents, between environments, and between agents and environments. The simulation then examines the output that results from the activation of the processes based on different system inputs.

An example of agent-based modeling is the work of Kalick and Hamilton (1986) on dating choices. Correlational studies have suggested that individuals tend to date those who are of equal attractiveness to themselves, often referred to as "attractiveness matching." Very attractive people tend to date very attractive people, moderately attractive people tend to date moderately attractive people, and unattractive people tend to date unattractive people. In traditional research, the correlation between a person's attractiveness and that of his or her partner is usually about 0.50. Some psychologists have hypothesized that people do not try to date potential partners who are more attractive than them because of fear of rejection and because people are more comfortable around people of similar attractiveness. In contrast to this correlational research and such conclusions, experimental studies suggest that people prefer to date very attractive partners, no matter what their own level of attractiveness is. How can this disparity between the correlational and experimental research be resolved?

Kalick and Hamilton (1986) conducted a computer simulation in which they created a population of 1,000 agents (i.e., simulated people), representing 500 males and 500 females. The individuals were randomly assigned a level of attractiveness ranging from 1 to 10. The male and female agents interacted at random and based on the other's attractiveness, one extended a dating offer to the other. If both agents made a dating offer, they were classified as a couple and removed from the dating pool. This continued until all agents had been matched.

In one condition, a dating offer was extended if the partner's attractiveness was roughly similar to the agent's level of attractiveness. Under this algorithm, the simulation produced a correlation between attractiveness levels for the final group of couples of about 0.85, which was too high relative to what correlational studies find. In another simulation trial, the probability that an agent extended an offer to a partner who had an attractiveness rating of 10 was set at 1.0; to a partner who had an attractiveness rating of 9, the probability of extending an offer was set at 0.90; to a partner who had an attractiveness rating of 8, the probability of extending an offer was set at 0.80; to a partner who had an attractiveness rating of 7, the probability of extending an offer was set at 0.70; and so on. Thus, the more attractive the partner, the more likely it was that he or she would be extended a dating offer. Under this algorithm, the correlation between attractiveness levels for the final group of couples was about 0.50, which maps on well to the correlational evidence. In studying the simulation results, Kalick and Hamilton found that the most attractive agents tended to couple up early on, so that the average attractiveness of the dating pool declined over time. This, in turn, resulted in the attractiveness of the couples decreasing over time. It was only through this analysis of the process of interaction dynamics that the correlational and experimental evidence in other studies could be reconciled.

Agent-based modeling frameworks are popular in psychology, sociology, and economics (Gilbert & Abbott, 2005). They have been applied to such diverse domains in sociology as the diffusion of norms, innovations to voting, the clumping of local networks into larger social structures, and the evolution of social structures, to name a few (Cederman, 2005).

RESOURCES FOR CONDUCTING SIMULATIONS

The Suggested Readings at the end of this chapter provide some useful resources for learning more about simulations. There are a host of computer languages for use in simulations, with a prominent one being NetLogo, which can be found at *ccl.northwestern.edu/netlogo*. NetLogo is a free, reasonably user-friendly modeling environment. More advanced tools include MASON by Luke, Cioffi-Revilla, Panait, and Sullivan (2004; see *cs.gmu.edu/~eclab/projects/mason*), and Repast (North & Macal, 2005; see *repast.sourceforge.net*).

SUMMARY AND CONCLUDING COMMENTS

A simulation is the imitative representation of the functioning of one system or process by means of the functioning of another. The original system that one seeks to simulate is called the *criterion system*, and the imitative representation of that system is called the *simulation*. Simulations can be an effective tool for theory construction, especially for problems focused on criterion systems that are complex and dynamic. Simulations differ from laboratory experiments in that simulations generally allow participants to control the flow of events (whereas in laboratory studies, this is often controlled by the experimenter), and noise and irrelevant factors are embraced rather than eliminated. Whereas laboratory experimentation generally concentrates on a limited number of independent and dependent variables, simulations generally include a much greater number of variables. That said, we should note that laboratory experiments and research simulations can also be integrated, so that a given simulation has participants randomly assigned to different treatment and comparison groups.

Research simulations serve two basic functions: (1) to build and clarify theories, and (2) to test theories. Among the different types of simulations are pedagogical versus research simulations, descriptive versus analytic simulations, real-time versus compressed-time versus expanded-time simulations, deterministic versus nondeterministic simulations, free versus experimental simulations, all-machine versus person–machine simulations, macro- versus microsimulations, and simulations varying in the content domain that is of interest.

One way in which simulations contribute to the theory construction process is by forcing the theorist to engage in a focused analysis of the criterion system in directions that might not otherwise have occurred. A second way that simulations contribute to theory development is through the conduct of virtual experiments. In this approach, a theory is translated into a computational model, and this computational model is subjected to initial validation by applying the model to hypothetical scenarios. Agent-based modeling is an extension of virtual experiments that also provides perspectives on theory.

The use of simulations for theory development is not without critics. Some scientists believe that simulations are too removed from reality and make too many simplifying assumptions. Simulations (especially those that do not involve human participation), the argument goes, simply cannot capture adequately the dynamic elements of how humans interact with their environment. The result is theories that are either too simple or, by virtue of their underlying mathematics and computer code, too complex. There are at least two counterarguments. First, since tightly controlled experiments and surveys conducted via interviews and self-administered questionnaires are generally unable to capture adequately the dynamic elements of how humans interact with their environment, behavioral simulations provide unique value for developing and testing theories regarding human behavior. Second, simulations may in fact be closer to the real-world phenomena they seek to study than are survey questionnaires and laboratory experiments.

Think about a phenomenon of interest to you. How would you design a simulation to analyze factors that impinge on it? Conduct a careful criterion system analysis from the perspective of trying to create and implement a simulation. What fundamental assumptions are you making? What core constructs and processes do you need to capture? What are the dynamic interrelations among them? Can you conduct a virtual experiment to gain insights into your theory and refine or expand it? What starting values would you put in place and later vary? What variables would you unpack and how? What assumptions would you make? What new features might you add to the system? As you attempt to answer such questions, we suspect that the richness of your theory will increase.

SUGGESTED READINGS

Davis, J., Bingham, C., & Eisenhardt, K. (2007). Developing theory through simulation methods. *Academy of Management Review, 32,* 480–499.—A condensed discussion of a wide range of simulation strategies as applied to theory construction about organizations.

Gilbert, N., & Troitzsch, G. (2005). *Simulation for the social scientist.* New York: Open University Press.–A description of a range of simulations and simulation strategies used in the social sciences. This book contains many references to current simulations being conducted in the social sciences from a traditional simulation vantage point.

Smith, E., & Conrey, F. (2007). Agent based modeling: A new approach for theory building in social psychology. *Personality and Social Psychology Review, 11,* 87–104.—An introduction to agent based modeling perspectives in the social sciences.

Sun, R. (2006). *Cognition and multi-agent interaction: From cognitive modeling to social simulation.* New York: Cambridge University Press.—An introduction to agent-based modeling.

KEY TERMS

simulation (p. 238)
criterion system (p. 238)
pedagogical simulation (p. 239)
research simulation (p. 239)
all-machine simulations (p. 242)
person–machine simulations (p. 242)
physical analogue simulations (p. 242)
mathematical simulations (p. 242)
descriptive simulations (p. 242)
analytic simulations (p. 242)
real-time simulations (p. 242)
compressed-time simulations (p. 242)
deterministic simulations (p. 242)
nondeterministic simulations (p. 242)
experimental simulations (p. 243)
free simulations (p. 243)
macrosimulations (p. 243)
microsimulations (p. 243)
content-oriented simulations (p. 243)
virtual environments (p. 250)
avatars (p. 250)
virtual experiments (p. 250)
agent-based modeling (p. 251)

EXERCISES

Exercises to Reinforce Concepts

1. What major functions do simulations serve?
2. There are a variety of different types of simulations. Identify and describe six of these.
3. For what principal reasons do scientists use simulations?
4. What is a criterion system? Provide three examples. Select one of the examples and identify the key features of this system that you think would need to be represented in any simulation designed to study that criterion system.

Exercises to Apply Concepts

1. Find an example of a simulation in the research literature. Describe it in detail. Propose a future simulation based on this one that will advance the theory underlying the simulation.
2. How would you go about developing a simulation of friendship formation? Of communication in your work environment? Of selecting a college or college major? Choose one of these topics, or another one of your choice, and describe the basic structure of a simulation that you would set up to help you derive a theory about the chosen phenomenon.

10

Grounded and Emergent Theory

People will occasionally stumble over the truth, but most of the time they will pick themselves up and continue on.

—WINSTON CHURCHILL (1940)

The most popular methods of scientific analysis in the social sciences are based on confirmatory frameworks whereby a scientist begins with a well-articulated theory and then subjects that theory to empirical test. The theory is derived from common sense, knowledge of previous research, and the type of logical and creative processes described in Chapter 4. An alternative approach to theory construction is called *grounded theory*. This approach emphasizes an approach of letting theory emerge from data rather than using data to test theory. In traditional social science, the researcher enters a research situation with an a priori theory, and the purpose of data collection is to "confirm" or "disconfirm" that theory; hence the phrase *confirmatory approach to science*. In grounded theory, data are not used to test an a priori theory. Rather, data are used to evolve a theory. Typically, the data are collected by qualitative methods that may include observation, analysis of archival records, structured and unstructured interviews, and focus groups, to name a few.

There are areas in the social sciences that tend to rely on *emergent theory orientations* but that do not fall into the formal camp of grounded theory. The grounded theory approach has its roots in sociology, and there is a specialized jargon and methodology that has built up around it. Although we will refer to many aspects of grounded theory in this chapter, we consider theorizing more from the perspective of anthropology, where emergent theorizing has been a cornerstone of the discipline for over 100 years. Readers with formal anthropological training will be surprised at times by some of the grounded theory jargon we use, and readers with more formal training in grounded theory will see us sometimes deviate from prescribed practice. But the spirit of both approaches should come through, and the suggested readings at the end of the chapter can serve as the basis for further study and elaboration.

We begin this chapter by providing an overview of grounded and emergent theorizing. We then discuss the frequently asserted stereotype that associates confirmatory or quantitatively oriented approaches to science with positivism and grounded/emergent approaches with constructivism. We argue that, in practice, epistemology is not strongly linked to the use of these frameworks. We next discuss how grounded emergent theorists frame problems and questions, and the role of literature reviews in this process. We then describe six different types of data upon which grounded emergent theorists often rely. This overview conveys a sense of the richness of the data sources that are used when constructing a theory. These include archival records, direct observation, structured and unstructured interviews, focus groups, virtual ethnographies, and directive qualitative methods. Next, we discuss key processes in theory construction, including memo writing, theoretical sampling, and the analysis and coding of data. This leads to a discussion of grounded emergent theory in terms of process analyses and the application of rhetorical theory to argument development. Finally, we discuss information-processing biases that are important for theorists to overcome when developing theory from data.

GROUNDED AND EMERGENT THEORY: AN OVERVIEW

Grounded theory is associated with the classic work of Glaser and Strauss. The first application of grounded theory was published in the book *Awareness of Dying* (Glaser & Strauss, 1965). The book *The Discovery of Grounded Theory* (Glaser & Strauss, 1967) was written 2 years later to make explicit the methods that had been used in this research. The book became the basis for grounded theory approaches in the 1970s and 1980s. Glaser and Strauss were heavily influenced by *symbolic interactionism*, a general theory of human behavior put forth by George Herbert Mead (1932) and later articulated by the noted sociologist Herbert Blumer (1969; see Chapter 11). Symbolic interactionism is based on three premises: (1) that people act toward things based on the meanings of those things to them, (2) that meaning is derived from social interactions (i.e., the meaning of objects emerges socially through our interactions with others), and (3) that meaning is the result of an interpretive process used by people to deal with the stimuli that they encounter. According to Glaser and Strauss, an important part of theory construction is discovering the meanings that different objects have to people and how their interactions are impacted by and define these meanings. This is best accomplished through intensive qualitative work based on direct contact with, and immersion in, the social world of those being studied.

Early writings on grounded theory emphasized that researchers were to set aside, as much as possible, preconceived ideas that they have about the phenomenon of interest and instead let relevant concepts and relationships emerge from rich qualitative data. In later years some grounded theorists have maintained this orientation, whereas others have encouraged the use of prior knowledge and cognitive heuristics to help explore the nature of meanings (Glaser, 1992).

Grounded theory construction typically begins by framing the research problem and initial research questions in general terms. Data collection then commences in which the researcher pursues data that are rich, substantial, and relevant to the problem at hand. As a reasonable amount of data starts to accumulate, the grounded theorist undertakes initial coding and interpretation of the data to gain a sense of the core concepts and the meanings of those concepts. A variety of coding strategies can be pursued, but a common strategy is one of comparing incidents (which we describe later). Throughout this activity, the theorist engages in a practice called *memoing* that involves writing ideas and conceptual notes "in the margins." These memos are consulted at a later time, when the theorist begins to move toward integrating categories, specifying relationships, and delimiting a theory. At each point in the process, the theorist considers the possibility of collecting additional data in ways that complement, expand, or better inform the emerging conceptual structure. This process is referred to as *theoretical sampling*. It is meant to ensure that the relevant conceptual universe has been thoroughly explored. All of the above eventually leads to the writing of the theory. In grounded theory, the tasks of coding, comparing, memoing, integrating concepts, and theoretical sampling are key processes.

In contrast to grounded theory, emergent theory approaches are not tied to symbolic interactionism, but they overlap with the grounded theory approach, more or less. These approaches emphasize description, understanding, and explanation, again relying heavily on qualitative data. According to this perspective, before one can theorize, one must first be able to *describe* the people, events, activities, meanings, contexts, environment, and culture of interest. As one gains a reasonable knowledge base along these lines, one seeks to *understand* what the different objects and events that have been described mean and represent, not only to the actors, but also to the scientist. Armed with such understanding, the theorist then seeks to *explain* why people behave or think as they do.

Imagine, for example, traveling to a remote area of South America and encountering a group of people who have been isolated from the rest of the world. In explaining the activities of this group, you might first write copious notes that describe the different people, events, activities, and contexts that you observe over an extended period of time. You might also learn to speak their language and gather as much information as you can about their history, by, for example, asking elders or examining relevant documents or historical residues. As you build your descriptive accounts, you try to understand what the different events, activities, and contexts mean to the people—that is, how they go about interpreting and orienting themselves to their environment. From this, you then build explanations of their behaviors—why they do what they do. It is within the context of such description, understanding, and explanation that theory emerges.

POSITIVISM "VERSUS" CONSTRUCTIVISM

Although there are excellent expositions of grounded and emergent theory, some writers polarize the approaches relative to confirmatory approaches to science. More often than not, a dichotomy is espoused that associates the confirmatory approach with positivism and the grounded approach with constructivism. Simplistic and unrepresentative versions of positivism are then subjected to intense criticism. We believe that such characterizations are unfortunate. In our experience, science in practice rarely fits neatly into simple dichotomies. Rather, scientists blend approaches in different ways and to differing degrees for different problems. In addition, almost all scientific approaches have something positive to offer, as long as one does not carry them to an extreme.

As an example, a recent book on grounded theory characterized, in a value-laden way, positivist theorists as seeking causes, favoring deterministic explanations, and emphasizing generality and universality. By contrast, interpretive theory, which the author associated with grounded theory, "emphasizes understanding, allows for indeterminacy, and gives priority to patterns and connections rather than to linear reasoning." We do not even know what "linear reasoning" means let alone why it is bad to engage in it. Nor do we know of any scientist who is uninterested in understanding the phenomenon he or she studies. Every scientist we know desires to discover patterns and connections. We know of no scientist who does not give great thought to the generalizability of his or her theories. Finally, why is seeking to find the causes of something so bad? One does not always have to seek causes, but doing so in some contexts and for some problems can be productive.

In the final analysis, we believe that one should strive to create a diverse set of tools for one's theoretical toolbox and then use them in ways that help get the job done. We recognize that orientations to scientific research differ when one has an a priori theory and seeks to collect data to test it versus when one has no theory and desires to use data to construct a theory. In the end, the ultimate goal is to describe, predict, understand, and explain behavior in ways that help us make sense of our world and that allow us to derive benefits. Both the confirmatory and emergent approaches are complementary, not conflicting, means to these ends. Confirmatory approaches need not embrace positivism, nor must grounded approaches embrace constructivism.

FRAMING THE PROBLEM

A common first step in building a grounded theory is to specify the problem area that one wants to address, the question that one wants to answer, or the phenomena that one wants to understand and explain (Spiggle, 1994). This is typically done in general terms, because the framing of the problem and question might change as one gains more firsthand knowledge in the field. The spirit at this step and throughout most of the analysis is to avoid imposing preconceived ideas onto the problem. There are, of course, occasions when theorizing is pursued in a more targeted way, such as when an existing theory is only partially developed and would benefit from additional exploration and probing. As well, there are scenarios for which emergent theorists avoid framing a problem statement at all, instead intending only to obtain a sense of "what's going on." But more often than not, emergent theorists begin with a statement of the general problem area or questions they want to address.

Emergent theorists recognize that it is impossible to approach theory construction with a complete blank slate, a *tabula rasa*. By definition, scientists bring to theory construction a set of concepts and meanings that are learned from birth and that allow them to make sense of the world in which they grew up and are now experiencing. An American social scientist may have a different "lens" through which to view events than a French social scientist or a Japanese social scientist. But the spirit of the emergent theory approach is to try to set aside preconceived notions so as to be open to new concepts and relationships, all the while being cognizant of the "filters" that impact the interpretations that are ultimately imposed. Strauss and Corbin (1998) speak of the construct of "theoretical sensitivity," which refers to the researcher's capacity, based on his or her experience, to focus on constructs that are central to the phenomena of interest.

An important part of framing the problem is specifying the population that you intend to study. For example, one might seek to study alcohol use in inner-city Latino youths in the United States. Or one might seek to study the relationship between economic ideology and economic practices among the Mayan peoples of the western highlands of Guatemala. Or one might study the behavioral mechanisms surrounding the spread of HIV and AIDS in young gay men in the United States. In each case, the specification of the group about which one wants to theorize delimits the scope of theory. To be sure, the precise target population might change as one proceeds further into data collection and analysis, but some delimitation of the population of interest inevitably occurs at the outset.

THE ROLE OF PAST LITERATURE

Like other approaches to theory construction, the timing as to when one should access relevant background literature is subject to some controversy in grounded theory frameworks. Some theorists argue that one should approach a problem after being fully informed by the extant literature, whereas others believe that this only channels and biases one's thinking. For example, Glaser (1978) recommends reading widely, but not pursuing the literature most closely related to the problem being researched so as to avoid channeled thinking. Some grounded theorists note that researchers may not even know at the outset the literature that will later be relevant.

Despite this orientation, the most common view is that past literature is an important resource that needs to be consulted prior to data collection. To most scientists, grounded or otherwise, it only makes sense to have a good working knowledge of the relevant theories and past research as they approach a given problem area. Indeed, the process by which master's thesis proposals, dissertation proposals, and grant proposals are evaluated at most institutions demands a careful consideration of the extant literature before embarking on fieldwork or the collection of data. The key is to ensure that such knowledge does not channel one's thinking too narrowly.

In some ways, the existing literature is a form of emergent data, just as archival records, interviews, focus groups, and observations are. The aim of the theorist is to compare the existing literature to the theory that emerges from data. Just as other data are subjected to coding and memoing, so should past literature be approached in this way, as one evolves a theory from all the different informational sources.

Of course, no matter what the orientation of grounded theorists is to past literature, when they write their theories, they include a scholarly and comprehensive consideration of past research in their writings. The issue being discussed here is the timing of when to bring the extant literature to bear on the thinking of the researcher.

COLLECTING QUALITATIVE DATA

Given a problem statement, the grounded/emergent theorist collects data to provide perspectives on the phenomena of interest. The data sought are intended to cast a wide net, as one seeks to describe, understand, and explain the phenomena, broadly construed. The term *ethnography* is used in anthropology to refer loosely to a broad class of qualitative methods applied with the purpose of providing a detailed, in-depth description of everyday life and practice. Because the theory construction process itself is so strongly tied to the data one collects, we briefly characterize some of the major methods that comprise ethnographies to provide a sense of the diverse sources of data one might consult for constructing a theory.

Archival Records

An important source for understanding the historical roots and contexts of many phenomena are archival records. If one wants to study adolescent drug use in the United States, for example, it may be helpful to explore as background the different policies and approaches to drug use that have been used in past decades, how the lives of adolescents over past generations have been affected by drug use, and the ways in which society has viewed and sanctioned adolescent drug use. In the field of anthropology, ethnohistorical methods have evolved that allow scientists to build an understanding of groups and societies through the historical study of maps, music, paintings, photography, folklore, oral tradition, ecology, archeological materials, museum collections, language, and place names.

Archival data can be sought from many sources. Municipal and state governments often have a wealth of archival data about business transactions, real estate and land transactions, zoning, and court cases, all of which can be revealing about the past of a community. Newspapers, magazines, and other forms of media from the past also offer a wealth of information.

Good historical methodology requires that one not only gather historical information from multiple sources but also critically evaluate those sources. For example, no person alive today personally knew Leonardo da Vinci, so how can we know anything

BOX 10.1. Anthropology and the Ethnographic Tradition

Emergent theory has had a central place in the historical development of anthropology as a field. The "father" of anthropology is generally recognized as Franz Boas, a German-born academic who was a professor first at Clark University in 1889 and later at Columbia University in 1899. Boas (1897) conducted research among the Kwakiutl Indians in British Columbia, Canada, and in the process developed new conceptualizations on race and culture. Boas felt that it was important to study all aspects of culture and that theory construction should be deferred until a group had been described as completely as possible. His meticulous and detailed ethnographies of the Kwakiutl reflected this orientation. Based on his emphasis on comprehensiveness, Boas advocated the "four-field approach" to anthropology, which emphasized human evolution, archeology, language, and culture. Each of these areas has since become a major subfield in anthropology. Boas was a highly influential figure on the conduct of ethnographies and the collection of qualitative data as a basis for understanding human behavior.

George Murdock was another influential anthropologist throughout the 1900s who advocated combining qualitative and quantitative methodologies. Murdock focused on comparative analyses of different cultures. He compiled ethnographies from around the world and in 1949, in conjunction with other researchers at Yale University, he established the Human Relations Area Files. In 1954 Murdock published a list of every known culture, entitled the *Outline of World Cultures*. In 1957 he published the *World Ethnographic Sample*, a data set of 30 variables for each of 565 cultures. In 1969 he worked with Douglas White to publish a data set of 186 cultures with approximately 2,000 variables describing each one. The Human Relations Area Files, now maintained by an independent research organization, consists of large data sets and extensive ethnographies of over 365 different cultural groups from around the world. The ethnographies alone consist of over 800,000 pages of indexed information about different societies. An extensive indexing system allows the ethnographies to be readily accessed and searched for purposes of comparative analysis.

As one might imagine, Murdock's work has been controversial among anthropologists, in part, because of its combined quantitative and qualitative emphasis. However, one cannot deny his early impact on the field. Other notable anthropologists in the history of the field who embraced ethnographic analysis include Kroeber, Levi-Strauss, Malinowski, Mead, and Radcliffe-Brown, to name only a few.

about him today? We have access to diaries, letters that he wrote, portraits of him, and we know what other people have said about him. However, all of this must be put together and interpreted. When we examine a document, it is essential that we evaluate the credibility of the source and the context in which the source is making his or her observations. Historians must consider when the source materials were produced, the conditions under which those materials were produced, the intentions that motivated the source, and the reliability of the source. They also must consider the broader historical context in which the documents were produced—the events that preceded and followed the creation of the documents—because the significance of any document depends as much on what comes after it as what came before it.

With the availability of high-speed computers and noninvasive scanners, archival materials are becoming increasingly available and accessible to researchers. Primary materials from the past are a rich source of ideas for theory building about current-day issues, and you will undoubtedly benefit from a study of historical and ethnohistorical methods and principles.

Direct Observation

Another rich source of qualitative data is direct observation. This takes many forms, including observing people at a single point in time to the classic method of participant observation. Sometimes people know they are being observed as part of a study and other times they do not. Sometimes the observer is physically present; sometimes the "observations" are recorded by an external device, such as a camera or a recorder, without the observer being present. Sometimes only the behavioral traces or behavioral residues that people leave behind are "observed" (e.g., studies that analyze the contents of people's garbage). Sometimes the scientist becomes a full-fledged member of the group being observed, indistinguishable from other group members. Other times the scientist tries to "blend in" as much as possible, but stands out in a recognizable way, such as the North American anthropologist who lives among the Maya Indians of Guatemala to try to better understand their customs and day-to-day activities. Scientists may observe people for only a short time or continuously for extended periods of time. Researchers may select a specific location at a specific time and observe all that happens at that time in that setting, or they may follow people to different locations to observe their behavior.

Crucial decisions with respect to observational methods include determining whom to observe, in what contexts to observe them, when to observe them, how to observe them, how often and how long to observe them, and in what form to make a record of those observations.

Observation can take many forms, and there is a degree of creativity in choosing the form of observation to pursue. For example, an academic social psychologist interested in understanding social influence processes took time off from his position at a major university and sought employment as a car salesman, an insurance salesman, a charity fund-raiser, an intern in an advertising company, and a retail salesperson—all with the idea of learning about influence strategies as used in these professions. A number of academic researchers in the field of consumer behavior traveled in mobile homes as a group for almost a year across the United States to many diverse locations, making extensive observations about their own and others' shopping experiences. Each night they would meet as a group and discuss what had transpired during the day, with the idea of formulating a comprehensive theory of consumer behavior. Neither of these observational enterprises is typical of formal observation methods, yet both resulted in insightful accounts of the phenomena the researchers were studying.

Structured and Unstructured Interviews and Surveys

Interviews can be structured or unstructured. Interviews are unstructured to the extent that there is no detailed script for asking questions, independent of the answers a research participant provides. Instead, the line of questioning is at the discretion of the interviewer and takes different directions, depending on the answers provided. By contrast, structured interviews have an a priori set of questions that is asked in a given sequence. The questions are formulated by the scientist ahead of time, so as to directly address the issues of interest, and the order of questions is minimally affected by the answers that are provided. Qualitative interviews typically record a person's responses verbatim, but occasionally the respondent is asked to choose between a small number of preestablished response categories.

Some interviews are completely structured whereas others are completely unstructured, but more often than not, there is a balance between the two in qualitative research. It is not uncommon for the first part of an interview to be unstructured and then to turn to a more structured format in the second part of the interview. A synthesis of structured and unstructured interviews creates the *semistructured* interview. Semistructured interviews have a range of topics and questions to be asked, but allow for new questions to be raised spontaneously by the interviewer, depending on what the interviewee says.

One popular type of interview is called a *life history*. It is a somewhat unstructured interview that is designed to elicit a sense of the life of the person being interviewed. The idea is that by understanding the life histories of key members of the target population, one will have a better understanding of the contexts in which behaviors are performed.

Life histories have different structures; a format that we find useful has several facets. After a set of introductory comments that establishes the focus for telling a life story, the respondents are asked to think about their life in terms of its "main chapters" and to tell us about those chapters, in whatever way they want. This part of the interview is unstructured. Next they are asked to identify a few key events that stand out in their life and to elaborate on each of them (e.g., what happened, how it happened, when it happened, with whom it happened, and what was special about it). After this discussion, respondents are asked to identify and talk about an event that defined the high point of their lives, an event that defined the low point of their lives, an event that was a turning point in their lives, their earliest memory, their most important childhood event, their most important event during adolescence, and their most important event as an adult.

Each event is thoroughly explored. Respondents are then asked to describe the single greatest challenge in their lives and then the biggest influences on their lives in terms of other people. Next, respondents are asked to consider their future and to describe it first in terms of hopes and dreams and then in terms of fears. Respondents are then asked to describe their values, political orientations, morals, and the spiritual side of their lives and how these have changed over time. The interview concludes with a final exploration of the respondents' lives in general as well as a solicitation for any other things we should know about to understand their lives. By the conclusion of the interview, which may take several hours, we have a rich sense of the individual's life.

A comparable type of interview to the life history focuses on *labor histories*, but in this case, the line of questioning focuses on jobs and economic matters rather than life more generally. In addition to life and labor histories, unstructured and semistructured interviews typically are used to focus more directly on the phenomena of interest to the theorist.

Dick (1990) describes a structured interviewing method called *convergent interviewing* in which two experienced theorists interview two informants, with one theorist interviewing one informant and the other theorist interviewing the other informant, each asking the same questions. The two theorists then compare notes, identifying common themes and points of agreement and disagreement between the two informants. The interview schedule is then revised for the next pair of interviews, so as to include probe questions to seek exceptions to the previous agreements as well as to seek explanations for the disagreements. The interview schedule is dynamic and shifts from pair to pair, as provocative issues are identified by the two different theorists.

There are many types of specialized interviewing strategies that have evolved in different disciplines. For example, *laddering* is a technique that has its roots in psychology and that has become popular in advertising and the analysis of consumer behavior (Veludo-de-Oliveira, Ikeda, & Campomar, 2006). It involves in-depth, one-on-one interviewing in which the focus is on eliciting basic personal values underlying the choices people make. The respondent is led through a sequence of questions, each time involving probes focused on "Why is this important to you?" and "What does it mean to you?" For example, a graduate student might be asked why he or she prefers one school over another, to which the answer might be because it is more affordable. The next question might be "Why is a school being affordable important to you?" Every answer the respondent gives is pushed further back to a more basic value by continually asking "Why is that important" to each response that is given.

Although interviews are the mainstay of a great deal of qualitative research, it is becoming more common for grounded and emergent theorists to pursue mixed-method forms of analysis, as discussed shortly.

Focus Groups

Another popular tool for qualitative researchers is the focus group. Questions are posed to a small group of individuals, usually 7–12 in number, in a group setting, and the group members provide their reactions and perspectives for all members to hear. For example, a focus group might be conducted with a group of students to explore how they view teachers, and a separate focus group of teachers might be conducted to explore how they view students. Like interviews, focus groups are conducted along a range from unstructured to structured. Participants are selected because they share certain characteristics that relate to the topic at hand. Focus groups are conducted by a moderator, who creates a permissive and nurturing environment that encourages different points of view. Multiple focus groups are often conducted on the same topic, with the composition of the groups sometimes varied strategically. The interactions are usually video- or audiotaped, and written transcriptions are derived from these recordings.

Focus groups can take many forms beyond the traditional focus group format. Some examples include (1) a two-way focus group, where one focus group observes another focus group and discusses the observed interactions and conclusions; (2) a "dueling moderator" focus group, where two moderators take opposite sides on an issue; (3) a participant-moderated focus group, where one of the participants is asked to assume the role of the moderator; (4) a teleconference focus group, where the participants are linked via teleconferencing; and (5) an online focus group, where computers and Internet connections are the main source of communication.

Focus groups can be useful in that the ideas and thoughtful expressions of one member of the group can serve as a catalyst for the ideas and thoughtful expressions of other group members. However, at the same time, the public forum may inhibit responses and frankness. With online focus groups, members can remain anonymous at the cost of losing some richness in communications. A dominant personality or influential member also can undermine a focus group. Sometimes the conversations in a focus group can become sidetracked and off point, in which case the moderator needs to keep participants "on task." Despite its weaknesses, the focus group can be a rich source of qualitative data.

Virtual Ethnographies

Virtual ethnographies are relatively new and extend the traditional notion of an ethnography to technologically mediated interactions in virtual settings and virtual communities. The physical boundaries of a virtual community can span the entire world. Just as ethnographers immerse themselves in the world of a target community defined in physical terms, the ethnographer studying virtual communities tries to do the same.

Virtual communities can be based on memberships in news groups, mailing lists, e-mail patterns, participation in virtual conferences, multisite users of a web page, or memberships or participation in chat rooms. Basic ethnographic questions about a virtual community include how many people belong to it, how long it has existed, how it defines itself, what its focus is, and who belongs to it. Many virtual communities are defined around topics; the nature and scope of these topics need to be documented. Often there is a special vocabulary or jargon that evolves in virtual communities, as well as a set of rules for communication, referred to as "netiquette" (Mason, 1996).

A researcher can collect virtual ethnographic data by participant observation (e.g., by being a member of that virtual community) and saving e-mails and/or listserve messages sent to, and received from, community members. One also can conduct electronic surveys and in-depth interviews of individuals, usually over the Internet, as well as download and analyze graphics and structures of web pages.

There are virtual communities on an incredibly diverse range of topics. As the influence of the Internet and virtual technologies continues to grow, this will be an additional source of data and information about many social-science-related phenomena. One of our anthropological colleagues, for example, was surprised to recently discover web pages for many of the villages she has studied for years in remote areas of Central America.

Directive Qualitative Methods

Some qualitative methods assign research participants tasks to perform and then use the data produced in the context of those tasks to construct a theory. An example of this approach is the metaphor elicitation technique (MET) developed by Zaltman for advertising research (Zaltman & Coulter, 1995). This method focuses on a product (broadly construed as a "topic") and asks participants to collect, prior to the main interview, photographs and/or pictures (from magazines, books, newspapers, or other sources) that indicate what the assigned topic means to them. Upon completing the task, a 10-step process is pursued using the pictures and images the person brings to the interview. Step 1 is called *storytelling* and involves having respondents tell how each picture is related to the topic. Step 2 focuses on *missed issues and images*, in which respondents describe any issue they were unable to find a picture about and to describe a picture that would capture that "missed issue." In Step 3, respondents perform a *sorting task*, where they are asked to sort the pictures into meaningful piles and to provide a label for each pile. There are no restrictions on the number of piles or the number of pictures in each pile. Step 4 is *construct elicitation*, wherein respondents articulate the differences between pairs of piles. In addition, the laddering technique discussed earlier is applied. In Step 5 respondents identify *the most representative image* and elaborate on why it was their choice. In Step 6 respondents characterize *the opposite image*; that is, they describe pictures that might typify the opposite of the task they were given. Step 7, called *sensory imaging*, asks respondents to use other sensory modalities to represent the topic. In Step 8 respondents work with the interviewer to create a *mental map*, based on everything that has been discussed in previous steps. This is a graphical device that positions each key construct identified in the previous steps in a separate rectangle or box, and then connects the boxes with lines to show the linkages (or lack of linkages) between them. In Step 9 respondents create a *summary image*, in which they make a montage of their images, putting them together so as to express issues that are important to them. Finally, in Step 10, called *making a vignette*, respondents create a vignette or story that communicates important issues related to the topic.

In sum, there are many methods for collecting qualitative data that can yield rich accounts of the phenomena in which a theorist is interested. A detailed discussion of these methods is the topic of a research methods book, not a theory construction book. Our intent is merely to provide you with a sense of the types of qualitative data with which grounded or emergent theorists work in the context of formulating their theories. Even the above is an inadequate representation, as there are many other forms of qualitative data that we have not mentioned (e.g., video recordings, collections of folklore and oral traditions, case studies).

Mixed-Methods Research

There has been debate in the social sciences about which method of data collection is best, the qualitative or quantitative method. Qualitative methods are associated with the data collection strategies discussed above, whereas quantitative methods are associated with surveys, closed-ended questionnaires, rating scales, and statistical analyses. It is becoming increasingly recognized that both approaches have strengths, so research that uses both approaches in complementary ways is becoming more common. Mixed-methods research is typically defined as research in which investigators mix or combine quantitative and qualitative methods of data collection (Tashakkori & Teddlie, 2002). From a theory construction standpoint, we believe that the use of multiple methods and multiple frameworks offers the best opportunities for generating new and creative ideas.

MEMO WRITING

As noted earlier, memo writing occurs at various phases of the data collection and theory construction process. Researchers write down field notes to themselves about ideas and insights they have "on the spot," which they then consult when analyzing their data, or when they formally posit their theory. One type of memo is called a *code memo*, which is a note relevant to the creation or coding of categories. A *theoretical memo*, by contrast, focuses on theoretical propositions linking categories or variables. *Operational memos* contain directions about the evolving research design and data collection strategies. Memo writing occurs in the field during data collection and also during data analysis.

Memo writing does not have to take the form of formal notes. For example, Zaltman (Zaltman & Coulter, 1995) sometimes writes memos in the form of vignettes that he thinks capture the kinds of metaphors that apply to his participants, or he draws images that graphically capture important dynamics that he observes.

Memos have a central role in grounded theory frameworks because they are often a crucial link between the data and the writing of the theory. They are a core part of the process of theory construction.

THEORETICAL SAMPLING

Grounded theorists emphasize the importance of theoretical sampling in the context of constructing theory from data. This is purposive sampling that often occurs after initial data have been collected and preliminarily analyzed. The purpose of theoretical sampling is to strategically increase the diversity of one's sample with the idea that this diversity will provide new information that will help one better appreciate and define the constructs and propositions that are evolving. The motivation behind theoretical sampling is not to obtain representativness, but rather to seek out new information that provides perspectives on the boundaries and nature of concepts and relationships between them (Charmaz, 2006).

Some grounded and emergent theorists emphasize analyzing qualitative data so as to understand the "three mosts"—that is, most of the thinking of most people most of the time (Zaltman & Coulter, 1995). This approach focuses attention on such matters as how many people mention certain constructs or certain themes. Theoretical sampling augments this orientation by encouraging the seeking out of new constructs, even if they are atypical, and exploring the limits of one's conceptions.

Related to the concept of theoretical sampling is *theoretical saturation*. This term refers to the decision to stop data collection because more data will not add anything new to the theory that has been, or is being, created.

The notions of theoretical sampling and theoretical saturation underscore another feature of grounded theory approaches: that data collection and data analysis often are realized simultaneously. It is only through the continual interplay between data collection and data analysis that one can make judgments about theoretical saturation and the need for additional sampling.

ANALYZING AND CODING DATA

The heart of grounded and emergent theory construction occurs at the level of data analysis. It is here that the scientist combines the insights gained during the act of data collection, the insights gained from reading past literatures, the field notes, and the information contained in the data to derive a theory. There is no single, correct way to abstract theory from qualitative data. Some theorists work solely at the level of narratives derived from a careful review of the data, others theorists rely on formal data coding, and others do both. Some researchers prefer a "top-down" approach, starting with general themes and then focusing on increasingly concrete representations within those themes. Other researchers prefer a "bottom-up" approach, starting with the concrete activities that people perform and then deriving more general themes from them. Some researchers prefer to embark on the process by avoiding, as much as possible, the imposition of an a priori framework on data interpretation, whereas other researchers apply a simple, flexible framework to start the analysis. Still other theorists impose a detailed framework, but always with a willingness to amend, adapt, augment, and drop categories and concepts.

An Example from Anthropology

Here, we provide an example of one approach to deriving theory from the coding of data that will make the above concrete, recognizing that it is only one of many different strategies that could be pursued. In this example, the researcher, Professor Liliana Goldin, is an anthropologist studying the impact of globalization on Mayan families and culture in the highlands of Western Guatemala. The particular research we discuss is a study analyzing the impact of maquila factories on Mayan families (Goldin, 2009).

The western highlands of Guatemala are a largely rural area, with townships and a few small cities located throughout. The traditional means of subsistence is through agriculture, in which a family grows crops on a small parcel of land, lives off a portion of the crops, and barters or sells the remaining crops at market. Mayan inheritance traditions are such that land is passed to all the children in the family, with the land parcel divided equally among them. With a high birthrate, this makes smaller plots of land available to each successive generation. The use of pesticides and certain types of fertilizers has made land productive for the Maya in the short run, but it also has undermined the fertility of the soil in the long run. On top of this, throughout history, the federal government has seized much of the most productive land of the Maya. Coupled with other political and natural events (e.g., genocide during civil wars, devastating earthquakes), many Maya have had to turn to means of survival other than agriculture.

With the rapid advent of globalization and the extremely low wages paid to Mayan Indians, maquila factories (also called export processing factories) have begun to appear in the highlands. These are foreign-owned factories that specialize in using unskilled labor to prepare exports for sale and use in other countries. For example, parts of clothing are sent from the United States to Guatemala, where the factory workers sew them together. The assembled clothes are then sent back to the United States, tax free, and sold to the American public.

The factories tend to hire young adolescent girls, who work long hours under oppressive conditions, typically for about $6 a day. The workers are thankful for the opportunity to earn a steady wage, although turnover is high, about 50% every 6 months. Many workers are fired (for not being productive enough, for doing poor quality work, for complaining, or for attempting to organize), and others quit because of the hardships of employment or for personal reasons.

The presence of the factories in communities is having a profound impact, not only economically, but also on the Mayan culture. For example, whereas an adolescent female previously spent time helping her parents in the field and around the house, this form of labor has changed. The girl who works in the factory now is a source of steady income and contributes to the household in new ways. Often the money she earns is used to pay for the education of boys, to buy food, and to help with general household expenses. Usually the girl is given a small discretionary amount of money to spend on herself.

A certain sense of independence is felt by the adolescent girl that comes with maquila work. The gender dynamics in the household are changing as a result of the factories and this, in turn, is affecting other facets of Mayan life.

Professor Goldin has conducted research in the Mayan communities for over 20 years. She lived in the townships with local families for 2 years and has returned to the area every year for extended stays. She is fluent in both Spanish and the Mayan language spoken in the area. Over the years, she gained considerable knowledge of the people, the culture, and the many changes that they have experienced. She also has researched their history through archival analyses, examining records dating to the invasion of the Spaniards in the 16th century.

The data for her study of the impact of the maquila factories include archival analyses; focus groups with maquila workers and nonworkers; interviews (including life and labor histories) of workers, nonworkers, and their families; discussions with owners; physical observations of the factories (though access is restricted by the factory owners and management); maps of the region and the locations of the factories on those maps; photographic accounts; and formal quantitative surveys. As an example, we consider a strategy she used to analyze interview data. For pedagogical reasons, we present a simplified account of the coding/analysis process.

Goldin started with a typology of the general areas that she wanted to explore. These included, for example, the effects of the maquilas on economic and labor practices, economic ideology, wealth and resources, gender dynamics, health, and family dynamics. She began by reading through all of the interviews so as to gain a general sense of their content and to see if any initial themes, propositions, or “story lines” jumped out. She then read each interview in earnest, placing a color-coded tag next to any segment that mentioned or dealt with economic issues, a different colored tag next to any segment that dealt with gender dynamics, and so on for each category of her typology. A given segment could be classified into more than one category. We call these level 1 categories.

Goldin defined a “segment” as any important thought or expression that the respondent made about a given topic. Such segments could be several paragraphs in length, a single paragraph in length, a sentence in length, or even a partial sentence. A key decision by a theorist analyzing qualitative data is specifying how a segment will be defined and how to identify the beginning and end of a segment. Some theorists define segments on a word-by-word basis in order to pursue a formal linguistic analysis of the data. Goldin defined a segment as a meaningful thought about a topic.

In the course of classifying segments into level 1 categories, additional areas on which she should focus suggested themselves, thereby creating new level 1 categories. For example, in her early studies with the Maya, it was serendipitously discovered that religious conversions from Catholicism to Protestantism were associated with the type of economic strategies that people tended to pursue. Further exploration of this result revealed a complex and fascinating story of the interplay between religious ideology and economic practice. A level 1 category on religion was therefore added to the framework.

Next, Goldin focused on a specific category 1 topic, say gender dynamics, and collected together all those tagged segments that focused on that topic. The segments for this topic were reread and a typology of themes within the gender dynamic topic area was developed. In other words, a set of level 2 categories was created within each level 1 category. The level 2 categories were derived, in part, using what grounded theorists call the *method of segment contrasts* or, more simply, *the method of contrasts* (Strauss & Corbin, 1998, also use the term *constant comparison* to refer to this method). In the method of contrasts, the theorist compares one segment with another segment, noting to him- or herself what is common and what is different about the two segments. This process is repeated many times, comparing different pairs of segments (in principle, all possible pairs of segments) until a conceptual framework for a typology starts to emerge, based on the many observed communalities and differences. During the process, notes are taken about the similarities and differences between segments and their possible implications. At some point, a formal typology of the level 2 themes within a level 1 category evolves. Goldin carefully reviewed her notes and elaborated and clarified the typology accordingly. She then assigned descriptive labels to the different categories or themes, a process that grounded theorists refer to as *naming*, and abstracted the major ideas being expressed. With the tentative typology in place, Goldin then repeated the process of reading each segment within the given level 1 category, but instead of comparing one segment to another, she now compared each segment to the taxonomic structure to determine which level 2 category fit it. Called the *method of classification*, this process led to further enhancements and refinements of the typology. Finally, the different segments were color-coded and tagged as to which level 2 categories they represented, thereby completing the level 2 categorization process. This process was repeated for each level 1 category.

The above strategy for creating level 2 categories is next repeated to create level 3 categories. A given level 2 category is selected and all segments that are tagged as members of the category are collected together. After an initial reading of all segments, the method of contrasts and the method of classification are applied, together with memo

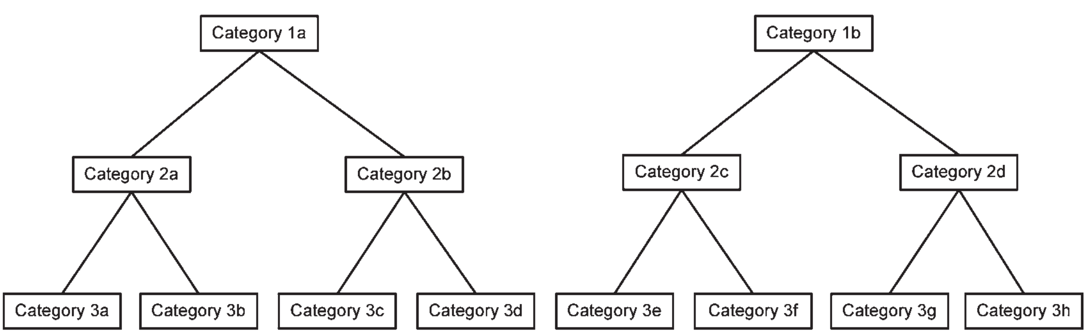


FIGURE 10.1. Tree diagram of category levels.

writing, to produce a typology of level 3 categories within each level 2 category. Figure 10.1 presents a tree diagram to reflect the resulting structure.

The process can be repeated yet again for the level 3 categories to create a fourth level of categories, and so on. At some point, the theorist makes the decision that further layering is beyond the scope of what he or she is trying to accomplish.

Once these steps are complete, theorists typically step back from the more micro-level analyses and think about the interconnections between the level 1, level 2, and level 3 categories more generally, in light of all the information processed. The idea is to adopt a "big picture view" and to examine the different phenomena as an interconnected system rather than focusing on its component parts. Indeed, knowing that this type of "big picture" analysis eventually will take place, researchers typically include notes and thoughts about the larger dynamics during the process of memo writing at the earlier stages.

Computer-Assisted Analysis

We described the above process referring to color-coded tags assigned to segments to flag the different categories of the segment at each category level. In point of fact, not only did Goldin do this, but she also created computer codes for each category (corresponding to each type of tag) for use in a computer program. Specifically, a given interview was broken down into segments, and each segment was then assigned a level 1 category code, a level 2 category code, and a level 3 category code. Multiple codes could be assigned if a segment fell into more than one category. This process was repeated for every interview and then the data for all interviews were merged into a single data file. The computer program was then used to sort through the segments to isolate almost instantly all segments associated with a given category at a given level. For example, when Goldin wanted to examine all segments focused on spousal interactions (a level 2 category) reflecting gender dynamics (a level 1 category), the computer sorted through the database and showed only these segments, with a respondent identifier for each one. In precomputer days, this process was pursued rather clumsily by sorting through index cards with one segment written on each card.

For the method of contrasts, the computer can display all possible pairs of segments at any given category level or it can display a random sample of the segment pairs or some systematically defined subset of pairs. This is important for several reasons. First, sometimes the sheer amount of information is overwhelming when pursuing the method of contrasts. For example, suppose there are 40 different segments to compare. In this case, there are 780 combinations of two segments to compare, which is a great deal of information to process. If 10 interviews are conducted and each interviewee mentions 10 segments in a given category, then there are 100 segments to compare. In this case, there are 4,950 unique combinations of two segments. It simply is not feasible for the human mind to process this amount of information. By forming all possible pairwise combinations of segments and then selecting a random sample of these using a computer, the task of systematically comparing pairs of segments becomes more manageable.

In sum, computers permit flexible and efficient accessing and pairing of segments. Instead of fumbling with index cards, pages of text, and tags, or instead of relying on memory, segments can be grouped, organized, and processed in diverse ways with a few keystrokes on a keyboard. This can facilitate theory development. Some of the more popular qualitative-based software programs include AnSWR, Atlas.ti, C-I-SAID, Decision Explorer, Ethnograph, N6, Nudist, and NVivo.

Defining and Manipulating Segments

In the preceding example, Goldin defined a "segment" as any important thought or expression that the respondent made about a given topic. Segments can be conceptualized using different criteria. Some researchers conceptualize segments broadly and others conceptualize them narrowly. Some researchers think about segments in terms of semantic meanings, others in terms of their affective content, others in terms of the processes they reflect, others in terms of who the actors are, and still others, at whom the action is being directed. Theorists do not necessarily choose between these different schemes. Rather, they can use all of them. Thus, multiple codes can be assigned to a segment, with some codes focusing on the meaning of the segment, some focusing on processes that are occurring in the segment, some focusing on the affective content of the segment, and so on. With the help of computers, one can access segments almost instantly using any given segment conceptualization that has been coded.

For example, one might have the computer focus on the affective content of all segments across all people interviewed and show first the text of all segments that are expressions of positive affect and then all segments that are expressions of negative affect. This could be pursued across all segments or, perhaps, focusing only on segments within a given level 1 category (e.g., those segments pertaining to gender dynamics). Or it could be pursued only for segments within a specified level 2 category (e.g., within gender dynamics, focusing on spousal interactions). Or it could be pursued only for specified level 3 categories.

To make matters more complex, the accessing of segments also can vary as a function of individual differences of the people interviewed. For example, the theorist might request that the computer show all segments expressing positive affect in the area of gender dynamics for male participants. This could then be repeated for female participants, with the theorist then making gender comparisons.

Guiding Questions and Coding Families

When thinking about defining and characterizing segments, some grounded theorists encourage analysts to use orienting questions to help them think about possible codes to assign and how to define segments. For example, Glaser (1978) described 18 "coding families" that researchers can use to develop codes and segments. Strauss (Strauss &

Corbin, 1998) described a coding family of "Six Cs" that stresses causes, consequences, and boundary conditions. Strauss and Corbin (1998) describe a set of diverse heuristics, much like those described in Chapter 4, as a way of exploring the different interpretations and conceptualizations one might apply to data.

Although numerous guiding questions and orientations have been suggested to help direct an analyst, there also are grounded theorists who believe that one should avoid such questions and coding families and simply let matters emerge from the data. You need to think about the approach that best resonates with you.

Open Coding, Axial Coding, and Selective Coding

Expositions of grounded theory offer a framework for coding data that is similar to the approach used by Goldin, but somewhat different. Distinctions are made between three types of coding: open coding, axial coding, and selective coding (Strauss & Corbin, 1998). *Open coding* involves developing initial categories for the words, sentences, phrases, or paragraphs under consideration. *Axial coding* involves imposing a coding scheme onto the categories from open coding that identifies connections between the categories. Strauss and Corbin draw the analogy of "putting an axis through the data" to connect the various categories identified in open coding. The connections between categories can come in the form of (1) conceptualizing the categories as causally related, (2) conceptualizing the categories as different aspects of a common dimension, and/or (3) conceptualizing the categories as part of a process, among others. *Selective coding* involves identification of the central themes and integrating the open and axial categories accordingly. The term *selective* is used because the analyst concentrates on a subset of core categories and connections around which to build the theory.

In sum, a great deal of thought is required regarding how to code qualitative data for later review and synthesis. Theorists can exercise considerable creativity in the ways in which they define and manipulate segments. Qualitative analysis is time consuming and difficult, but it offers many rewards in terms of theory construction.

THE STATISTICAL EXPLORATION OF RELATIONSHIPS

Grounded and emergent theorists are interested in exploring relationships between concepts, categories, and variables in the context of the qualitative data they collect. This process sometimes can be assisted by the use of statistical methods. Some grounded theorists oppose the use of statistical methods in the theory construction process, complaining that it removes them too far from the data and encourages simplistic thinking. Sometimes this argument is valid and sometimes not. It is beyond the scope of this book to describe the many ways in which statistical analysis of ethnographic data can foster effective theory construction. We encourage readers to keep an open mind rather than embrace general condemnations of statistical or quantitative analysis.

PROCESS ANALYSIS IN EMERGENT THEORIZING

As noted earlier, many emergent theorists emphasize process-oriented perspectives in their theorizing. Surprisingly little, however, has been written on how to think in terms of processes on a concrete, practical level. We discussed such matters briefly in Chapter 4 and consider them again in Chapter 11, but some observations, as derived from the grounded/emergent theory, literature are offered here.

Webster's dictionary defines a process as "a systematic series of actions directed to some end" (e.g., the process of homogenizing milk) and as "a continuous action, operation, or series of changes taking place in a definite manner" (e.g., the process of decay). The essence of these definitions involves the notions of action and change—that is, describing actions that lead entities from Point A to a different Point B. Charmaz (2006) suggests that the following questions be asked when thinking in terms of process:

1. How does the person act while engaged in this process (i.e., what are the actions and states that describe the process)?
2. How does this process develop?
3. How does the person claim to think and feel in this process?
4. What might the person's observed behavior indicate or signify while engaged in the process?
5. When, why, and how does the process change?
6. What are the consequences of the process?

A useful tool for thinking about processes is the process map, a graphical device, much like a flow chart, that summarizes the flow of a process. Constructing process maps helps the theorist see the sequence of events, the branches that can occur in a process, and how all the different events in a process are related to one another. It can be a powerful heuristic device for theory construction because the sheer act of creating one often leads to insights that might not otherwise be gained.

Here are the major steps involved in creating a process map: (1) Determine the start and termination points of the process, (2) list the different events in the process, (3) sequence the events, and then (4) think about the logical relations among the events. The process map is then drawn using a conventional symbol notation, as shown in Figure 10.2. In constructing process maps, the theorist must think about how detailed the events in the process map should be. The more specific and detailed the events are, the more intricate the mapping. Variants of process maps used in qualitative research include flow charts, decision trees, and kinship maps.

We consider additional strategies for thinking about change and processes in Chapter 11. As we have stated in earlier chapters, it is rare to find emergent theories that focus exclusively on process. Rather, there is a blending of process- and variable-oriented thinking that draws upon many of the tools for explanation that we have discussed thus far.

MOVING TO THEORETICAL STATEMENTS: USING PRINCIPLES OF RHETORIC

After formal analysis of data, memos, and field notes, the researcher sets about the task of describing his or her theory. Scientists have different ways of doing so. Some scientists describe the relevant variables within the theory and the relationships between them, developing the logic and past empirical support for those relationships. These scientists might use path diagrams to help them present their theories (see Chapter 7). Other scientists present the theory in the form of propositions, conclusions, or theses, and then weave a set of arguments in support of those statements. The arguments are logically derived from common sense, past theory, empirical data from past research, and/or the formal ethnographic research used to construct the theory. All of these approaches to theory description have merit.

Emergent theorists tend to frame their theories using propositions and then develop supporting arguments. Given such an orientation, it is useful to consider perspectives from the field of rhetoric, which describes strategies for forming reasoned and convincing arguments in support of conclusions and theses. We do so now, deferring until Chapter 12 a discussion of strategies for writing about grounded and emergent theories. The focus here is on building effective argument structures in support of propositions. We recognize that there can be more to theories than presenting propositions and arguments in support of those propositions. However, a good portion of many theories adopts this general strategy; hence it is instructive to consider principles from rhetoric.

Deduction, Induction, and Abduction

Rhetoricians distinguish three means of persuasion: (1) *logos*, which are strategies that rely on logic and emphasize induction and/or deduction; (2) *pathos*, which are strategies that rely on emotions and emotional reactions; and (3) *ethos*, which are strategies that rely on attributions of expertise, trustworthiness, or charisma. Science, ideally, invokes logos.

Two types of argument structures based on logos are *induction* and *deduction*. Induction is a type of reasoning wherein one provides a limited number of examples and, from these examples, infers a general rule or principle. In other words, one moves from specifics to the general. Consider this example of inductive reasoning: John was not admitted to college *X*, Barbara was not admitted to college *X*, and Joan was not admitted to college *X*; therefore, college *X* is a difficult school to which to gain admittance. The important feature of this strategy is to ensure that the examples used are representative of the general case. Thus, when utilizing an argument based on induction, one needs to make a convincing case that one can validly infer the general rule from the specific instances under consideration and that the specific instances are not exceptions. This standard should be applied, for example, when a grounded or emergent theorist is providing exemplars of a theme or exemplars to support an argument or proposition.

Deduction is a type of reasoning wherein one asserts a general principle that all

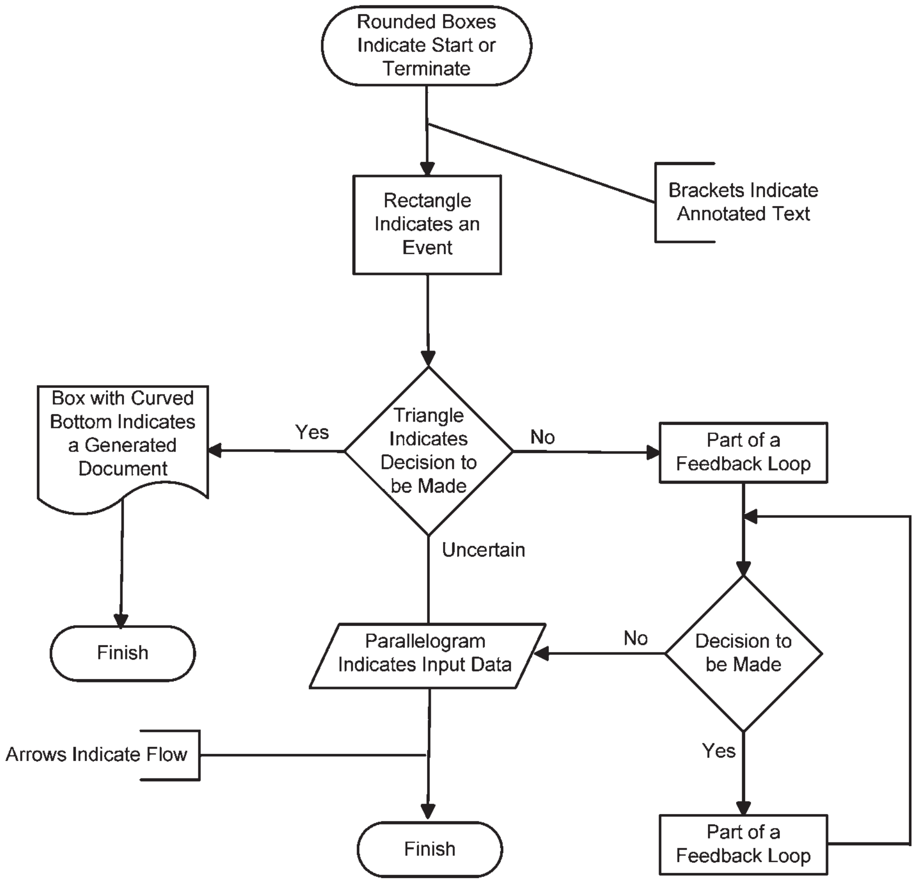


FIGURE 10.2. Conventional symbols for a process map.

agree is true and then argues that the case at hand is an instance of it. For example, we might all agree that immigrating to a new country for the first time is stressful in the early stages of the immigration process. Because John is a new immigrant, we conclude that he is experiencing stress in his life. The important feature of this logical strategy is to ensure that the example fits into, or is part of, the larger category to which the rule applies and which everyone agrees is true.

Both induction and deduction are reasonable strategies of argumentation, and scientists make use of one, the other, or both a great deal. It sometimes is helpful to analyze your theoretical arguments to determine if you are relying on induction or deduction. If you are relying on induction, have you made a strong case that the exemplars are representative and that the general rule reasonably follows from them? If you are using deduction, have you made a strong case that the instance you are considering is indeed a member of the more general class or group to which consensual rule applies? And is the consensual rule indeed consensual and widely accepted?

The philosopher Charles Peirce (Martin, 1979) suggested another form of inference, called *abduction*: that is, inference to the best explanation. This is a method of reasoning in which one chooses the hypothesis that would best explain the relevant evidence. Abductive reasoning starts from a set of accepted facts and infers their most likely explanation. It stands in opposition to Popper's falsification principles (1963). According to Peirce, there are hundreds of possible explanations for a phenomenon, and it is not possible to empirically falsify every possibility. Instead, we "abduct" only those explanations that are more plausible. Abduction is not a form of symbolic logic but instead a form of critical thinking. It is controversial in the field of rhetoric and philosophy because most treatments of inference emphasize the rules of formal logic.

Toulmin's Model of Argumentation

The British philosopher Stephen Toulmin (1958) suggested a theory of reasoning in which argumentation is said to have six interrelated parts. The first part is the *claim*, which is the conclusion, proposition, or thesis that the scientist seeks to establish. The second part is the *data*, which are the empirical "facts" to which the scientist appeals in support of the thesis. The third part is the *warrant*, which are statements that help link the data to the claim. For example, if the claim is that "standardized tests like the SAT [Scholastic Aptitude Test] are biased against minorities," a piece of data in support of this might be that Latinos who have the same grade-point average in school as European Americans tend to score lower on the SAT than the European Americans. A "warrant" for this claim–data link is that a person's grade-point average in school is a reasonable indicator of the underlying abilities that the SAT is intended to measure. A fourth part of the argument is the *backing*, which are statements and supporting evidence that certify the warrant. Backings are statements that do not support the main point directly, but that support the warrants that link the data to the claim. In our SAT example, they would be evidence for the assertion that a person's grade-point average in school is a reasonable indicator of the underlying abilities that the SAT is intended to measure. Backings are important because if the warrant is invalid, then the validity of the argument is undermined, which weakens the claim. A fifth part of the argument is the *rebuttal*, in which counterarguments or positions opposed to the argument are considered and refuted. The sixth part of the argument, the *qualifiers*, articulates the degree of certainty in the claim and describes exceptions to the claims and the limits of its generalizability.

There may be multiple pieces of data in support of a claim, multiple warrants linking a given datum to the claim, multiple backings for a given warrant, and a complex structure of rebuttals and qualifiers. We find it useful to apply Toulmin's (1958) model to theoretical analyses as a way to better isolate the nature and strengths of the lines of reasoning being developed. For a given claim, what are the different pieces of data being invoked to support it? For each datum, what are the warrants? For each warrant, what are the backings? Have counterarguments been taken into account? How have these been refuted? What qualifiers need to be applied and taken into account? Application of the Toulmin framework reveals both the strengths and weaknesses of an argument structure and can be useful for scientists as a tool to help refine their theories.

Weak Arguments

Theories of rhetoric identify weak forms of argumentation or weak appeals that appear relevant to a claim but that really are not relevant. These represent "traps" of which scientists should be cautious as they describe the support for their theses, conclusions or theoretical propositions. Knowledge of such traps also can help you critically evaluate theories that you read, as sometimes theorists invoke them as a form of weak argumentation. We describe 16 common examples, but there are many more, in practice:

1. *Argumentum ad hominem* (argument toward the man). This strategy praises or criticizes people who make a claim, rather than discussing the claim itself. It occurs when scientists make personal attacks against a critic, as during debates about the merits of a theory.
2. *Argumentum ad populum* (argument to the people). This strategy asserts that because most people believe that a claim is true, then the claim must be true. Popular acceptance of a claim does not validate it. Just ask Galileo or Copernicus.
3. *Argumentum ad traditio* (appeal to tradition). This strategy asserts that a claim must be true because it is part of tradition. A variant of this is that because a claim was true in the past, it must be true in the future. Such logic does not always hold.
4. *Argumentum ad verecundium* (appeal to improper authority). This strategy invokes someone with expertise in another area that is not really appropriate for the claim at hand. For example, one might invoke the endorsement of a claim by a Nobel prize-winning economist about the appropriateness of a reading program for elementary-school-age children.
5. *Argumentum ad misericordiam* (argument from pity). This strategy makes an emotional appeal to try to convince someone to accept a claim based on emotion, independent of the logic underlying it.
6. *Petitio principii* (begging the question or circular reasoning). This strategy uses a claim to support an argument underlying that claim, which is circular. For example, a reinforcement theorist might argue that "a mother reinforces a child's behavior and therefore maintains the behavior by paying attention to the child every time the child performs the behavior." How does the theorist know that giving attention to the child is reinforcing? "Because it maintains the behavior." This is circular reasoning.
7. *Dicto simpliciter* (hasty generalization). This strategy improperly uses inductive reasoning when there are too few samples to prove a point. One is too quick to form a generalization from specific instances.
8. *Non causa pro causa* (not the cause for a cause). This strategy invokes a false cause of an event for the real cause. Consider the claim "Drinking wine in moderation, say, a glass a day, reduces the risk of heart disease." This claim asserts a false cause. More specifically, there are socioeconomic differences in wine-drinking behavior, with higher-class individuals being more likely to drink a glass of wine a day than lower-class

individuals (who are more likely to drink beer). Better access to health care reduces the risk of heart disease, and upper-class individuals have better access to health care than lower-class individuals. Asserting that drinking a glass of wine a day has beneficial health effects is a position that is *non causa pro causa* because the real source of the effect is social class, not drinking wine. If social class is held constant, there is no association between wine consumption and the risk of heart disease.

9. *Ignorantio elenchi* (irrelevant conclusion). This strategy uses an argument for one claim and redirects it to inappropriately support a different claim. A variant is to change the subject or divert the argument from the real question at issue.

10. *Non sequitur* (it does not follow). This strategy invokes a claim that does not follow from the previous statements or arguments. Logicians often use the term in reference to formal syllogistic errors. For example, arguing that marijuana is a gateway drug to heroin use because almost all heroin users started on marijuana is a *non sequitur* (we discuss why this is so later in the chapter).

11. *Argumentum ad ignorantium* (argument from ignorance). This strategy appeals to a lack of information to prove a point, or arguing that since the claim cannot be disproved, it must be true. For example, one cannot prove that ghosts do not exist, therefore they must exist.

12. *Argumentum ad speculum* (hypothesis contrary to fact). This strategy tries to prove something by using hypothetical examples that are not applicable.

13. *Either–or fallacy*. This strategy creates false dichotomies, arguing that there are only two choices when actually there are more.

14. *Equivocation*. This strategy uses a word or term in a different way than the author used it in the original claim or argument.

15. *Stacking the deck*. This strategy involves ignoring examples that run counter to a claim, focusing only on examples that support it.

16. *Argument from the negative*. This strategy states that because one position is untenable, the opposite position must be true. Both could be in error.

Principles of rhetoric have much to offer scientists in their efforts to develop effective theories, as scientists consider the evidence and various arguments for or against theoretical propositions. We encourage you to explore this discipline in your further readings. Carefully analyze your propositions and the argument structures you evolve in support of them. Consider if you are invoking induction, deduction, and/or abduction and make sure you do so appropriately. Map out the claims, the data, the warrants, the backings, and consider rebuttals and qualifiers, per the Toulmin model. Finally, review your arguments and make sure you that are not falling prey to using one or more of the weak argument forms described earlier.

SUMMARY AND CONCLUDING COMMENTS

A major approach to theory construction is that of grounded or emergent theorizing. Whereas the dominant method of scientific analysis in many areas of the social sciences is based on confirmatory frameworks (i.e., using data to test theory), emergent theory emphasizes a process of letting theory emerge from data. Grounded theory is associated with the classic work of Glaser and Strauss (1967), with early approaches emphasizing its link to the theory of symbolic interactionism. However, many emergent theory frameworks are not tied to symbolic interactionism, and, indeed, the discipline of anthropology has a longstanding tradition in emergent theorizing that is distinct from grounded theory. The emphasis in anthropology is on description, understanding, and explanation.

Grounded/emergent theory construction typically begins by framing the research problem. Data collection then commences, usually taking the form of qualitative data. The methods used include archival research, direct observation, structured and unstructured interviews, focus groups, and virtual ethnographies, to name a few. The data sought are intended to cast a wide net, as one seeks to describe, understand, and explain the phenomena of interest, broadly construed. The term *ethnography* is used in anthropology to refer loosely to a broad class of qualitative methods applied with the purpose of providing a detailed, in-depth description of everyday life and practice.

As a reasonable amount of data start to accumulate, the grounded theorist undertakes initial coding of the data to gain a sense of the core concepts and the meanings that seem to be operative. The coding and analysis of data often involve defining segments. Typically, a "segment" is any important thought or expression that a research participant makes about a given topic. However, segments can be defined using a variety of criteria. A common strategy for analyzing the meaning of segments is the method of contrasts, which involves comparing segments and creating categories from this comparison process. At some point, a formal typology is evolved, and then the researcher compares segments with the evolved taxonomic structure. This is called the method of classification.

Memo writing occurs throughout the research process; researchers record notes to themselves about ideas and insights they have "on the spot." Grounded theorists also emphasize the importance of theoretical sampling, which is a purposive type of sampling that often occurs after initial data have been collected and preliminarily analyzed. The purpose of theoretical sampling is to increase the diversity of one's sample, with the idea that this diversity will provide new information that will help one better appreciate and define the constructs and propositions that will constitute the theory.

After analysis of the data and memos, the scientist creates the formal theory. A common strategy used by emergent theorists is one of presenting the theory in the form of propositions, conclusions, or theses, and then elaborating a set of arguments in support of those propositions and conclusions. Theories of rhetoric and argumentation are useful in helping grounded/emergent theorists specify their theory in effective ways. Rhetoricians distinguish three types of arguments: logos, pathos, and ethos. Logos arguments rely on logic and emphasize induction and/or deduction; pathos arguments rely on emotions and emotional reactions; and ethos arguments rely on attributions of expertise, trustworthiness, or charisma. Science, ideally, invokes logos.

The British philosopher Stephen Toulmin (1958) suggested a six-part model that can be used to analyze argument structures. Theories of rhetoric also identify weak forms of argumentation that appear relevant to a claim but that really are not relevant. These represent "traps" of which scientists should be cautious as they describe the support for their theses, conclusions, or theoretical propositions. Examples include *argumentum ad hominem*, *argumentum ad populum*, *argumentum ad tradition*, *argumentum ad verecundium*, *argumentum ad misericordiam*, *petitio principii*, *dicto simpliciter*, *non causa pro causa*, *ignorantio elenchi*, *non sequiturs*, *argumentum ad ignorantium*, *argumentum ad speculum*, either–or fallacies, equivocation, stacking the deck, and argument from the negative.

In emergent theory construction, social scientists rely on their ability to observe others and derive theoretical insights from those observations. Psychologists have identified limitations to human information processing that can bias and shape the way in which we think about things (see Appendix A). These include limitations of human memory, the conditional symmetry bias, the tendency to see systematic events in random events, the compound probability bias, bias due to the availability of information in memory, order effects, and difficulties ignoring irrelevant information. Graduate training in disciplines that emphasize qualitative methods needs to educate students about these limitations and then train them to either eliminate them or to take the biases into account.

Grounded and emergent theorizing are useful and productive approaches to building theories. It is hard to argue with the idea that a powerful strategy for building a theory is to immerse yourself in as rich and meaningful a way as possible in the phenomena you are studying and the lives of the people directly involved with that phenomena. Some argue that one loses objectivity with such immersion, but, from the perspective of theory construction, it is unclear if this argument justifies abandoning such an obviously rich source of ideas. There are, of course, some areas where grounded theory approaches are limited. For example, researchers who study the neuroscience of attitudes with the goal of mapping electrical activity in the brain associated with drug addiction probably are not going to get very far in their quest by pursuing grounded theory approaches that interview people about why they use drugs and the meaning and place of drugs in their lives. Or are they? Perhaps the nature of electrical activity in the brain is tied to the way in which people construe and think about drugs, with different parts of the brain being activated in different ways as a function of these different ways of thinking about things. Keep an open mind as you think about ways to build your theory. Just maybe the approaches that your first reaction is to dismiss will turn out to be the basis for a novel insight.

SUGGESTED READINGS

Barber, R. J., & Berdan, F. (1998). *The emperor's mirror: Understanding cultures through primary sources*. Tucson: University of Arizona Press.—A discussion of strategies for archival data analysis.

Denzin, N., & Lincoln, Y. (2000). *Handbook of qualitative research*. Thousand Oaks, CA: Sage.—A major resource for a variety of qualitative methods.

Glaser, B. G. (1992). *Basics of grounded theory analysis: Emergence versus forcing*. Mill Valley, CA: Sociology Press.—A rebuttal of a grounded theory approach advocated by Strauss and Corbin that highlights points of controversy in the grounded theory framework.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.—A comprehensive and classic exposition of the grounded theory approach.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge, UK: Cambridge University Press.—A reader on basic biases identified by cognitive and social psychologists.

Kelle, U. (1996). *Computer-aided qualitative data analysis: Theory, methods and practice*. Thousand Oaks, CA: Sage.—A book of readings on computerized methods for analyzing qualitative data.

Locke, K. (2001). *Grounded theory in management research*. London: Sage.—A clear, short introduction to grounded theory.

Miles, M., & Huberman, M. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks, CA: Sage.—A classic on qualitative data analysis.

Strauss, A. L., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: Sage.—Description of practical strategies for implementing the grounded theory approach.

Strauss, A. L., & Corbin, J. (1997). *Grounded theory in practice*. Thousand Oaks, CA: Sage.—An edited book with 10 chapters illustrating grounded theory construction principles.

Tashakkori, A., & Teddlie, C. (2002). *Handbook of mixed methods: Social and behavioral research*. Thousand Oaks, CA: Sage.—A book of readings on mixed-methods approaches to the conduct of research.

van Eemeren, F. H., Grootendorst, R., & Henkemans, A. F. (2002). *Argumentation: Analysis, evaluation*. Mahwah, NJ: Erlbaum.—A concise introduction to theories and principles of argumentation.

KEY TERMS

grounded theory (p. 257)
emergent theory (p. 257)
ethnography (p. 261)
archival records (p. 261)
direct observation (p. 263)
structured and unstructured interviews (p. 264)
life history (p. 264)
labor history (p. 265)
convergent interviewing (p. 265)
laddering (p. 265)
focus groups (p. 265)
virtual ethnographies (p. 266)

code memo (p. 268)
memo writing (p. 268)
theoretical memo (p. 268)
operational memo (p. 268)
theoretical sampling (p. 269)
theoretical saturation (p. 269)
method of contrasts (p. 272)
method of classification (p. 272)
open coding (p. 275)
axial coding (p. 275)
deduction (p. 277)
induction (p. 277)
logos (p. 277)
pathos (p. 277)
ethos (p. 277)
abduction (p. 279)
Toulmin model of argumentation (p. 279)
claim (p. 279)
warrant (p. 279)
backing (p. 279)
argumentum ad hominem (p. 280)
argumentum ad populum (p. 280)
argumentum ad traditio (p. 280)
argumentum ad verecundium (p. 280)
argumentum ad misericordiam (p. 280)
petitio principii (p. 280)
dicto simpliciter (p. 280)
non causa pro causa (p. 280)
ignorantio elenchi (p. 281)
non sequitur (p. 281)
argumentum ad ignorantium (p. 281)
argumentum ad speculum (p. 281)
either–or fallacy (p. 281)
equivocation (p. 281)
stacking the deck (p. 281)
argument from the negative (p. 281)
sensory memory (p. 288)
short-term memory (p. 288)
long-term memory (p. 288)
magic number 7 plus or minus 2 (p. 288)
working memory (p. 289)
conditional symmetry bias (p. 289)
compound probability bias (p. 291)
availability bias (p. 292)
base-rate fallacy (p. 293)
confirmation bias (p. 294)
egocentric bias (p. 294)
false consensus bias (p. 294)
halo effects (p. 294)
illusory correlation (p. 294)
negativity bias (p. 294)
self-serving bias (p. 294)
unrealistic optimism (p. 294)
wishful thinking (p. 294)

EXERCISES

Exercises to Reinforce Concepts

1. How does an emergent theory approach differ from the confirmatory approach to science?

2. Provide an overview of the grounded theory approach to theory construction.
3. What are the key points that you should take into account when framing a problem in emergent theory construction?
4. Briefly characterize the major methods for qualitative data collection.
5. What is a life history interview?
6. Describe different kinds of focus groups.
7. What is a virtual ethnography?
8. What is memoing? Why is it important?
9. What is theoretical sampling? What is theoretical saturation? How are the two related?
10. Describe the process of creating multiple levels of categories when coding data.
11. What is the difference between the method of contrasts and the method of classification?
12. In what ways can computers assist in the processing of qualitative data?
13. What are the advantages and disadvantages of using family codes and guiding questions?
14. What are logos, pathos, and ethos?
15. What is the difference between induction, deduction, and abduction?
16. Describe Toulmin's model of argumentation and discuss why it is relevant for theory construction.
17. Identify and define 5 of the 16 types of weak arguments. Give an example of each.
18. What is the magic number 7 plus or minus 2, and what are its implications for constructing theories?
19. Give an example of the conditional symmetry bias.
20. Give an example of the compound probability bias.
21. How is availability of information to memory important for how a scientist builds a theory?

Exercises to Apply Concepts

1. Find an example in the research literature of a study that presents a theory based on either grounded or emergent theory approaches. Describe the theory

in depth and the methods the theorist used to evolve the theory. Analyze the theory using all the principles of rhetoric and argumentation discussed in this chapter.

2. It is not feasible to conduct a full ethnographic accounting of a phenomenon and then develop your own theory about it in the context of a one-semester course on theory construction. However, you can expose yourself to features of the approach on a small scale. For example, you might go to a church, a fraternity party, a cafeteria, a factory, a community center, a classroom, or some other setting and make observations about a phenomenon of interest to you. Use all of the principles discussed in this chapter to make your observations and then try to derive a small-scale theory based on these. Conduct relevant literature reviews. After building your theory, carefully analyze it in terms of the facets described in the preceding exercise.

3. In the first class I ever taught (Jaccard) as a new professor, I assigned a large class of 250 students the task of thinking of a social norm and then going out and breaking it during the ensuing week. The students were to write a report of how people reacted to it and, just as important, what their own reactions were while breaking the norm. Based on this experience, the students were to write an analysis of social norms and, in essence, build a mini-theory of social norms. I felt that the exercise combined many interesting features of a field study that would serve class discussion well. It was not until the next class, when the students reported some of the activities in which they had engaged, that I realized I had made a serious mistake, essentially giving students a valid excuse for acting strangely. This once-in-a-Jaccard-lifetime assignment was quite successful in the theoretical insights it yielded as well as the recognition by students of how powerful norms are in American society. We are *not* assigning this exercise to you here.

Appendix 10A

THE LIMITS OF INFORMATION PROCESSING

In emergent theory construction, social scientists rely on their ability to observe others and derive theoretical insights from those observations. Psychologists have identified limitations to human information processing that can bias and shape the ways in which we think about things. In this appendix, we describe a number of these processing limitations, so that you can be aware of the ways in which they might color your view of the world as you approach the theory construction process. These limitations may cause you to think more narrowly than you want, make faulty inferences, or focus on concepts or ideas to the exclusion of others that are important. A good grounded/emergent theorist will be aware of these tendencies as he or she approaches field observations and do his or her best to avoid bias because of them. Indeed, this is true of any theorist who relies on observation or inference to form theoretical propositions.

THE MAGIC NUMBER 7 PLUS OR MINUS 2

One important class of limitations for making inferences is the limitation of human memory. Psychologists often think of human memory as being like a computer. The computer receives information from the environment (a keyboard), converts that information into a specialized code, the code is stored in immediate memory (random access memory; RAM), and then some is saved to a hard disk. Later, the information is retrieved, either from RAM or from the hard drive. Humans are exposed to information from the environment. They encode this information and then store it in temporary memory. Some of this information is passed to more permanent memory, which can then be retrieved for later use.

Psychologists distinguish three types of memory. *Sensory memory* stores all stimuli that register on the senses (e.g., touch, sight, hearing). Information is held in sensory memory only for a fraction of a second to 3 seconds. If the sensation does not draw the attention of the individual, the memory vanishes forever. If attention is drawn to it—that is, if we notice it—it is encoded and then passed into short-term memory, also called working memory.

Short-term memory is a temporary storage system, where a piece of information is stored and held for perhaps 20 seconds, after which it fades away and is lost. Some of the information in short-term memory is passed to *long-term memory*, which is a more permanent memory that can hold vast amounts of information. Short-term memory is not just a passive bin of storage for sensory input information. Rather, humans "move" memories from long-term memory into short-term memory and then manipulate the contents in complex ways. Short-term memory is a dynamic information-processing center, but it also is fleeting. It is for this reason that many psychologists prefer to call it *working memory*.

What is the capacity of working memory? In a classic paper, George Miller (1956) argued that humans can hold about 7 chunks of information in working memory. He refers to "the magic number 7 plus or minus 2" because across a wide range of studies and tasks, the evidence suggests people can simultaneously hold between 5 and 9 chunks of information in working memory. To be sure, sometimes a chunk of information can be composed of many organized bits of information (e.g., a chess master's vision of a chessboard with all the pieces organized in a well-known configuration), but the human mind has clear limitations to the number of chunks of information it can process in working memory at one time. Simon (1974) has suggested that the magic number 7 might actually be somewhat less, closer to 4 or 5.

This limitation is important because it suggests that scientists are restricted in the number of variables, concepts, and relationships to which they can attend at any given instant. For theoretical analyses that are complex, scientists must rely on recording information externally and then processing that information sequentially. There also is evidence that people do not necessarily process information in the ways they think they are processing it. For example, when making judgments, people tend to think they use more information than they actually do, and they tend to think they use it in more complex ways than they actually do (Wiggins, 1973). To us, all of this underscores the importance of using thinking aids to assist in processing large amounts of information when constructing a theory (e.g., aids such as a path diagram discussed in Chapter 7, a process map, or a systemized strategy of note recording). It also underscores the importance of the computer (and, in some cases, multivariate statistical methods) that can simultaneously process relationships between hundreds of pieces of information. The human mind, unassisted, is simply incapable of doing this.

The point we wish to stress is the importance of recognizing that we are limited in our abilities to process large amounts of information, and we need to approach our theorizing strategies and interpretation of data accordingly.

CONDITIONAL SYMMETRY BIAS

People often process probabilistic information, but they do not always do so in conformity to objective probability theory. An example is the *conditional symmetry bias*. You may have heard the assertion that marijuana is a gateway drug that leads to heroin use. As evidence, some cite the fact that the proportion of heroin users who have smoked marijuana is large. This proportion represents what statisticians call a conditional probability, symbolized as $p(M|H)$, where M is past use of marijuana and H represents using heroin. The symbol $p(M|H)$ is read as "the probability that a person has used marijuana, given that the person is a heroin user"; this conditional probability is usually quite large, near 0.95. But this in no way means that the reverse conditional probability, $p(H|M)$, is also large: $p(H|M)$ is the probability that someone is a heroin user given that the person has smoked marijuana. As it turns out, this probability is quite small (near .05). The conditional probabilities are not symmetrical (i.e., equal in value) even though many people assume they are. The tendency to assume equal conditional probabilities is called the conditional symmetry bias.

As another example, a survey of 74 CEOs working for Fortune 500 companies led a scientist to conclude recently that there may be a link between childhood pet ownership and future career success. About 95% of all the CEOs as children had possessed a dog, a cat, or both. The CEOs that were interviewed commented on how pet ownership had helped them to develop many of the positive character traits that made them good managers, such as responsibility, empathy, and respect for other beings. Note that the datum being reported is p(owned pet in childhood|CEO) = .95. Because this probability is large, the inference is made that the two variables are linked. But what we really want to focus on is p(CEO|owned pet as child): that is, given that you owned a pet as a child, what is the probability that you will become a CEO? This is very, very small. Here, again is the erroneous assumption of symmetry in the conditional probabilities.

We have found several cases of conditional symmetry bias in scientific reports and in scientific theories. When inferring relationships, one must be careful not to commit this logical fallacy. A correct inference of a relationship goes well beyond considering the information in the conditional probabilities $p(A|B)$ and/or $p(B|A)$. For details, consult any introductory book on probability theory. Although emergent theorists rarely formally calculate conditional probabilities, they do make inferences about relationships based upon observations they have made and notes they have taken. The inference of a relationship might be based on a subjective or intuitive representation of a conditional probability. For example, after interviewing a large number of heroin users, the scientist might make the observation that almost all of them started their drug habit using marijuana. The inference of marijuana as a gateway drug may then be formulated, even though we know that such an inference cannot be made from such data.

WHEN RANDOM IS NOT RANDOM

Random events happen all the time, but people sometimes find it difficult to recognize randomness. People impute meaning into random events and build theories around them as if they were systematic, usually because they do not understand the dynamics of randomness.

As an example, suppose we flip a coin four times. Which of the following sequences is more likely to occur:

Sequence *A*: H H H H
Sequence *B*: H T T H

Most people will say sequence *B*. The correct answer, however, is that they are equally likely to occur. Each has a probability of .0625 of occurring. This is obvious if we list the 16 different outcomes that could occur on the four flips:

H H H H
H H H T
H H T H
H H T T
H T H H
H T H T
H T T H
H T T T
T H H H
T H H T

T H T H
T H T T
T T H H
T T H T
T T T H
T T T T

Note that the probability of sequence *A* occurring is 1 out of 16, and 1/16 = .0625, and so is the probability of sequence *B*.

In a random process such as multiple coin flips, people think that the number of heads will equal the number of tails, and when this expectation is violated grossly, they tend to make attributions of nonrandomness that can be incorrect. In the example above, on how many trials in the four coin flips do we get an equal number of heads and tails? From the above listing, of the 16 possible outcomes, six have two heads and two tails. Thus, the probability of observing two heads and two tails is 6/16 = .375. This means that it is almost twice as likely that one will *not* observe an equal number of heads and tails in a given random sequence of four flips than it is that one will observe two heads and two tails. This is counterintuitive to most people.

There are many features of randomness that are nonintuitive and we need to be careful not to let our intuitions about randomness lead to attributions of systematic relations in what are essentially random events. Training in probability theory and statistics helps one to appreciate these nuances.

COMPOUND PROBABILITY BIAS

Another common bias is the *compound probability bias*. Suppose a woman has just started using birth control pills. Let *A* be the event of a minor side effect occurring, say, innocuous cramps, and suppose that the probability of this event is .80. Let *B* be the event of her boyfriend approving of the use of the birth control method and suppose that the probability of this event also is .80. What is the probability of both events occurring; that is, what is the probability that both the minor side effect and boyfriend approval will occur? Most people will say it is .80, but is this correct?

It turns out that the correct answer is .64. Specifically, in probability theory, if events *A* and *B* are independent (which is reasonable to assume in this case), then it can be shown that $p(A \text{ and } B) = p(A)p(B)$. That is, the probability of both events occurring equals the product of the two probabilities that each event will occur in isolation. The compound probability fallacy is that people tend to overestimate the joint occurrence of events, given their perceptions of the probabilities of each individual event.

Here is another example. Suppose you are in charge of a complex missile system whose successful functioning depends on each of 500 parts working correctly. Suppose that the probability of a part functioning correctly the first time it is used is .99. What is the overall probability that the system will work the first time it is used? Most people would say it is quite high. Assuming independence, however, the probability is $.99^{500}$, which equals about .01. The chances that the missile system will work the first time are about 1 in 100, or 1%! Just as laypeople are susceptible to the compound probability bias, so may the scientist be susceptible to such bias when conducting field observations and making judgments about the occurrence of joint events.

AVAILABILITY OF INFORMATION FROM MEMORY

When we are asked to make a judgment about something, we frequently search our memory and then formulate a judgment based on the information that is retrieved from memory. The information that influences our judgments most is the information that is easily retrieved from memory. The problem is that information that is most available in memory is not always the best information on which to base a judgment. What factors influence availability of information to memory? There are many. One factor is how recently we have been exposed to, or thought about, the information. For example, Higgins, Rholes, and Jones (1977) primed participants in an experiment by having them read lists of positive or negative attributes. Later that day, when they were asked to make attributions of an ambiguously described person in a different experiment, the research participants were more likely to use categories and words that had been primed earlier. Identical behaviors were interpreted as "self-confident" or "arrogant" depending upon whether positive or negative words had been used for the priming task in the previous experiment. Such "priming effects" can have obvious implications for judgments a scientist might make in the field because such judgments can be colored by the information to which the scientist has been exposed (i.e., primed with) prior to making the observations.

Unique events also stand out and hence, they are more likely to capture our attention and be encoded and passed to long-term memory or retrieved from long-term memory. Suppose we read to you the following list of names:

Steve Smith
Joan Gerber
Bill Clinton
Jeanne Weiss
John Glass
Martha Smith
George Bush
Glenn May
Michelle Gayle
Lilly Stanton

After we read the list aloud to you, we might ask you how many males were on the list. This is the task that Kahneman and Tversky (1972) used in their research. They read lists of names to people and then asked them to estimate the number of males and females on the list some time later, without forewarning people that they would be asked to make such a judgment. The correct answer for the above list is five males and five females, that is, half the list has male names. In the Kahneman and Tversky study, most people said the number of males was two or three. The research participants judgments were biased toward the number of famous males in the list (Bill Clinton and George Bush), because these names were more prominent and hence more "available" in memory.

ORDER EFFECTS

Suppose we describe the same person to two different individuals. The description provided to the first person is that the target individual is intelligent, sincere, honest, conceited, rude, and nervous.

The description provided to the second person is that the target individual is nervous, rude, conceited, honest, sincere, and intelligent. Note that the exact same information has been provided to each person, and only the order of it has been changed. For the first person, the positive descriptors come first followed by the negative descriptors (pro–con). For the second person, the order is reversed (con–pro). One would expect that the reactions to, and evaluations of, the described individual would be the same, on average, since exactly the same information has been presented. But this is not the case. When people are presented descriptors in the pro–con order, they tend to form more favorable impressions of the individual than when they are presented the descriptors in the con–pro order. The order in which we are presented or exposed to information can bias the judgments we make. Scientists are just as susceptible to these order effects as laypeople, and care must be taken to counteract this bias when making field observations and collecting and analyzing data.

IGNORING INFORMATION

Sometimes we are exposed to information, but we want to ignore it or not take it into account when making judgments. For example, in some court cases, jurors are told something by a lawyer, the opposing lawyer makes an objection, the objection is sustained, and then the jurors are instructed to ignore the information. But do they? Cialdini (2001) describes a study in which people were randomly selected from jury lists and then exposed to a recording of a trial about a woman who was injured in a car accident by a reckless driver. Jurors were randomly assigned to one of three conditions. In one condition, the juror heard the case and no information was provided about whether the driver had liability insurance. In a second condition, the juror heard the case, and it was revealed that the driver had liability insurance. In the third condition, the juror heard the case, it was revealed that the driver had liability insurance, and then the judge disallowed the information and asked the juror to ignore it. If people ignore the information, then judgments in the third condition should be the same, on average, as judgments in the second condition. The mean award by jurors for pain and suffering in the first condition was about $33,000, for jurors in the second condition it was about $37,000, and for jurors in the third condition it was about $46,000. Not only did the jurors fail to disregard the information, they apparently gave it more weight by being told to ignore it. Later studies revealed that jurors tended to give it more weight because they saw it as "privileged" information.

Just because we want to ignore information does not mean that we can indeed ignore it when forming judgments. When judging the type of soil in a plot of land for making planting decisions, farmers are told that they should ignore the moisture content of the soil samples. Yet invariably, the moisture content impacts their judgments about how much of the core components of the soil is present (e.g., the amount of sand particles, the amount of clay particles). And if we are told by an informant that information is special, we tend to give it more weight in our judgments, even if the information is not special (Cialdini, 2001). Such biases can color the judgments scientists make in the context of their field observations, and they need to be dealt with accordingly.

ADDITIONAL PROCESSING LIMITATIONS

Psychologists have identified a wide range of additional limitations that people bring to bear as they process information and make judgments, among them (1) the base-rate fallacy (the tendency to ignore base-rate information when making frequency judgments), (2) the confirmation bias (the tendency to seek information that will confirm one's expectations or beliefs), (3) the egocentric bias (the tendency to see oneself as the center of attention; the tendency to see oneself and one's activities as unique), (4) the false consensus bias (the tendency to overestimate the degree to which other people agree with oneself), (5) halo effects (the tendency to evaluate the specific attributes of an object based on one's overall evaluation of the object), (6) illusory correlation (the tendency to see two unrelated events as being related), (7) the negativity bias (the tendency to weigh negative information more heavily than positive information), (8) the self-serving bias (the tendency to attribute successful outcomes to self and failures to extenuating circumstances), (9) unrealistic optimism (the tendency to distort reality in a direction that enhances self-esteem and maintains personal efficacy and optimism), and (10) wishful thinking (the tendency to see desirable outcomes as more probable and undesirable outcomes as less probable). All of these can impact the judgments of the people who report about the world about them as well as the judgments of the scientist directly.

Graduate instructors, training in disciplines that emphasize qualitative methods, need to educate their students about the limitations of human information processing and then help train them to take them into account and, where possible, adopt strategies to overcome them. In our experience, these disciplines could benefit from multidisciplinary interactions with psychologists and decision theorists to better keep abreast of these limitations and methods for overcoming them. We encourage you to make learning and keeping up on this literature part of your training and professional growth as a scientist.

11

Historically Influential Systems of Thought

A hunch is creativity trying to tell you something.

—Frank Capra (1956)

There are many ways of thinking about the world, and we have tapped into only a few of them. In this chapter we briefly describe historically influential systems of thought that large numbers of social scientists have used to theorize about diverse phenomena. We also consider some lesser known but still influential thinking strategies that might be of use to you as you approach the theory construction process. We advocate what is sometimes known as meta-triangulation in theory construction, or the building of theories from the perspective of multiple paradigms (Lewis & Grimes, 1999). After introducing you to a dozen or so different systems of thought, we encourage you to think about the phenomena in which you are interested from each perspective. That is, think about the phenomena through the lenses of different thought systems and see what new ideas and insights result. To be sure, not all of the systems will "work" for you, but even if only a few do, you will be that much further ahead.

We do not describe the different systems in depth; doing so would require book-length manuscripts on each topic. Instead, we expose you to the spirit of each framework. As you become more familiar with the frameworks in the course of your studies, they may take on greater or lesser import in your scientific efforts. There are ardent critics of each framework, including some who would be appalled that we even mention a given framework. Similarly, some critics will feel that we have left out an important framework. Whatever the case, we believe that each framework has something of value to contribute.

The frameworks are an eclectic group that can be organized in multiple ways. We grouped together one set of frameworks under the general rubric of "grand theories" as discussed in sociology and anthropology. These include materialism, structuralism, functionalism, symbolic interactionism, and evolutionary perspectives, along with a critical commentary of these grand theories from the perspective of postmodernism. Next, we discuss frameworks that draw heavily on metaphors. These include neural networks and systems theory. We then consider frameworks that emphasize the analysis of change followed by two psychological frameworks. We conclude with the discussion of frameworks inspired by methodological innovations, namely multilevel modeling and person-centered theorizing.

Some of the frameworks are more popular in some disciplines than others. For example, materialism is more widely used to analyze phenomena in sociology and anthropology than in psychology. This should not, however, deter you from using a framework, no matter what your discipline. Also, some of the frameworks are dated in that their dominance in a given discipline has waned with time. This also should not deter you from using the framework as a way of possibly generating new perspectives and ideas about the phenomena you are studying.

GRAND THEORIES

In the fields of anthropology and sociology, distinctions often are made between grand theories and middle-range theories (Mills, 1959). Although the characterizations of these two approaches vary, grand theories are seen as comprehensive, inclusive theories of human behavior and society, whereas middle-range theories are more focused accounts of specific phenomena (but for an alternative characterization, see Alford, 1998). Most of the social sciences currently operate at the level of middle-range theories, but there are notable grand theories that theorists often draw upon in formulating their middle-range theories. The present section considers a selection of these grand theories: materialism, structuralism, functionalism, symbolic interactionism, and evolutionary perspectives. We also discuss postmodernism, though it certainly is not a grand theory in the tradition of the other frameworks. Rather, it is a critical commentary on some of the fundamental assumptions underlying the grand theories.

Materialism

Materialist theories emphasize the analysis of human behavior and social institutions from the perspective of the material aspects of society. The approach has its roots in the work of Karl Marx and Friedrich Engels. Although Marxist theory primarily has been used to motivate political ideologies and political analysis, it also has been used as a form of scientific analysis (Plattner, 1989). Marxist frameworks emphasize the concepts of history, ideology, and the analysis of inequalities in social power and wealth. When trying to understand human behavior, one examines the types of ideologies that people have held and how these ideologies are shaped by inequalities in wealth and the means of producing that wealth. A materialist pays careful attention to who creates the products that society values and how they go about doing so, how products are converted into wealth, and how the wealth is distributed among different segments of society.

According to Marx, economic inequalities often are structured along social lines that ultimately define socioeconomic classes. In capitalist societies, wage workers comprise one class and capitalists comprise another class. In materialistic models, class distinctions also are extended to other groups, such as men versus women and the elderly versus youth, primarily based on their control over the means of production. Materialism involves a careful analysis of the historical context in which the above relations emerge and change over time, as well as those aspects of ideology, production, and class that have remained stable. Also of importance is identifying the ideological means by which the members of the lower class have accepted their positions and by which members of the upper class justify their high status. These systems of beliefs, also called *hegemonies*, are an integral part of materialist analysis.

At the most general level, materialist frameworks highlight the central role of economic issues in impacting behavior and society. One of the authors (Jaccard) remembers early in his career explaining his research on teen pregnancy to a cultural anthropologist with a Marxist bent. He described how he analyzed the beliefs and attitudes that teens have about sexual activity, the peer pressures that teens experience, the emotions and affect surrounding adolescent sexual relationships, how issues of sexuality impact the self-concept of adolescents, and, finally, the important role of feelings of efficacy. The anthropologist politely listened for about half an hour, even though at times she looked bewildered by all the jargon being thrown about. She responded with a two-sentence reaction: “Jim, it’s class. It’s all about social and economic class.” Of course, there is more to adolescent pregnancy than class issues, but the comment started Jaccard thinking about the broader contexts of adolescent sexual risk behavior. Further reflection led to an analysis of the economic underpinnings of adolescent risk behavior and the class dynamics that are involved.

There are numerous examples of useful analyses that have adopted materialist perspectives. Sidney Mintz (1986) conducted a thorough study of sugar production in the Caribbean and highlighted the ways in which the introduction of this commodity changed world economic relations, generated new forms of labor, augmented the slave trade, and completely transformed Caribbean economies. The newly acquired tastes of the English ruling elites for sugar led to global needs for sugar and contributed to a forceful new dynamic in the world economy. In typical form of a materialist, Mintz focused on history, power relations, and the economy to provide a complex and integrated analysis of social, cultural, and economic processes surrounding sugar production.

In another insightful analysis, June Nash (1989), in her book *From Tank Town to High Tech: The Clash of Community and Industrial Cycles*, studied deindustrialization in the northeastern United States, documenting the impact on a town of the closing of a General Electric plant that built products for the military. She showed how the changes associated with the closure paralleled changes in the country and in the world more broadly and how these macro-level changes could be used to make sense of the local developments. Her analyses emphasized economic dynamics at the local, national, and world levels and their impact on the lives of individuals.

The materialist framework is not without its critics, but one cannot deny the potential centrality of the kinds of questions it poses when analyzing cultural and societal influences on behavior. Consider carefully how different goods related to the phenomena in which you are interested are produced, allocated, and valued in society. Pay attention to the sources of power, who controls them, and how that power is exercised and transferred. Think about potential conflicts between groups over the allocation of goods and the monopolization of sources of power. Think about how all of these factors permeate political ideologies, economic systems, religion, and educational systems and how these, in turn, impact behavior.

As you apply the above, be flexible in conceptualizing what constitutes "goods." For example, suppose you are studying abortion. How is abortion made available, allocated, and valued in society? Who are the sources of power with respect to abortions (e.g., hospitals, clinics, government, religious organizations, Planned Parenthood)? Who controls these organizations (e.g., owners of hospitals, owners of clinics, politicians, religious leaders)? How is control of these organizations exercised and passed on from one year or one generation to the next? What types of conflicts have occurred between which groups over the *Roe v. Wade* decision that made abortions widely available to women? How do political ideologies, economic systems, religion, and educational systems impact access to abortion? The materialist framework encourages you to think about your phenomena on a more societal level in the context of issues of power, class relations, and control of resources.

Structuralism

Structuralism is a system of thought that derives from linguistics and is most strongly associated with the writings of Claude Levi-Strauss. It is related to the theory of transformational-generative grammar developed by Noam Chomsky, which we use as an analogy for describing structuralism (though Chomsky based his analysis, in part, on the work of Levi-Strauss). According to Chomsky, underneath the surface structure of language is a "deep structure" that represents a finite set of organizing principles that serve as a universal linguistic blueprint for all languages. By carefully analyzing the surface structure of many different languages, one can isolate this blueprint. Chomsky argues that analyses of language at the surface structure makes them appear more diverse than they really are. At a deeper level, all languages have a small number of organizing principles, and the task of the linguist is to discover these.

Levi-Strauss, an anthropologist, approached the analysis of culture from this same perspective. Although cultures appear diverse on the surface, he argued that there is a core set of underlying organizing principles and structures. Structural analysis of a phenomenon focuses on trying to isolate the basic underlying structural principles associated with it. Levi-Strauss stressed the importance of both the conscious and the unconscious when considering the bases of human behavior.

Levi-Strauss conceptualized human thought in terms of the principle of binary opposites (good–evil, light–dark, tall–short); that is, he argued that people think in a binary or dialectical nature. Thoughts are expressed differently in cultures through transformations, but such transformations are constrained by the structural operations of the human mind. Structuralism focuses on discovering *how* people think rather than *what* people think. Levi-Strauss used the metaphor of code, emphasizing the importance of decoding surface structure events into their deeper meanings. In doing so, Levi-Strauss argued that the scientist should seek analysis of contrasts and opposition.

As an example of a structuralist perspective, Ortner (1974) conducted an analysis of why women have subservient roles and are judged as being inferior relative to men across cultures, worldwide. She argued that the disparity derived from oppositional ideologies that associated women with nature and men with culture. She used these opposing associations to articulate structuralist arguments in support of her thesis. Ortner's work has been controversial and her thesis has been refuted, but her work crystallized questions and new perspectives on gender and culture (e.g., MacCormack & Strathern, 1980). As another example, Goldin (1987) conducted a structural analysis of Spanish and Mayan assumptions about markets and exchange in Western Guatemala. She identified two contrasting interpretive frameworks used by the Spaniards and the Maya. Whereas the Spaniards preferred enclosed, elevated, artificially illuminated market environments in which to sell goods and the goods organized in accord with Spanish classifications, the Maya preferred open markets, with goods laid out on the ground, natural lighting, and a categorization system consistent with Mayan culture. Her analysis of these views highlighted the oppositional nature of the underlying logic of the two cultures, which is consistent with a structuralist perspective.

Structuralist thinking has its strong and weak points, but for purposes of this book, we highlight the following structuralist principles: (1) Think about your phenomena by finding the deeper structure underlying the seemingly diverse surface structure events surrounding it, (2) consider the potential role of both conscious and unconscious factors that might be operating, and (3) think about matters in binary or dialectical terms, focusing on opposites and contrasts (see Rychlak, 1994, for an interesting application of dialectical thought).

Functionalism

Functionalism has been an influential conceptual framework in sociology and anthropology. Notable functional theorists include Durkheim, Evans-Pritchard, Malinowski, Merton, Parsons, and Radcliffe-Brown. There are two major types of functionalist theory, one focused at the societal level (as typified by the work of Durkheim) and the other focused at the individual level (as typified by the work of Malinowski), although there is overlap between them. Some refer to the societal-based frameworks as *structural functionalism* and the individual-based frameworks simply as *functionalism*.

Every society has functional requirements for its survival. Functional analysis explores social institutions and segments of society in terms of the functions they serve to this end and conceptualizes society as a system of interdependent parts that tend toward equilibrium. A society is in equilibrium if there is no conflict, if people know what is expected of them, and if societal expectations are met. Parsons (1951, 1971) argued that equilibrium is attained through socialization processes and through social control—that is, sanctions imposed either informally through norms and peer pressure or through formal organizations, such as schools and prisons. Functionalists think of society as a collection of systems within systems. For example, Parsons discusses the personality system within the small-group system, the small-group system within the community system, the community system within society, and societal systems within the world system.

Merton (1968) distinguished between latent and manifest functions. Latent functions are consequences of a cultural or institutional action that are not explicitly intended or recognized by members of a society. Manifest functions are consequences that contribute to equilibrium and that are intended and recognized by societal members. An example of a latent function was described by Edwards (1979), who contrasted the "efficiency movement" with the "human relations movement" in the field of industrial relations. The former movement traditionally was thought to emphasize efficiency in organizational settings through strict control of the workers, whereas the latter movement was thought to downplay control in favor of a more humanistic orientation toward increasing productivity. Edwards's analysis showed that when examined at the latent level, the human relations movement exerted just as much control over workers as the efficiency movement, but in different and subtler ways.

Functionalism has fallen into disrepute in some areas of the social sciences. For example, the idea of a society functioning to achieve equilibrium has been criticized because so many segments in society seem to operate independently of other segments. Functionalism also has been criticized for being ahistorical, focusing on the functions that institutions serve at present rather than how they evolved over time and the historical forces that shaped them. Despite these limitations, we find it useful to sometimes think about a phenomenon or an entity in terms of the kinds of functions it serves, say, for the individual, the family, the school, the neighborhood, or the community. Instead of asking about the causes of something, ask what functions it serves. Think of the systems in which the phenomena are embedded and ask what functions those systems serve. As you articulate functions, think in terms of those that contribute to the betterment of the individual, community, and/or society. What are the manifest functions that are operating and what are the latent functions? Needs are important components of functional analyses (e.g., social needs, economic needs), because functions often are associated with the meeting of such needs. What are the needs that define a functional analysis for your topic area? Consistent with criticisms of functionalist approaches, think of whether there is conflict and lack of equilibrium in the organization, institution, or problem you are considering. Is the conflict intrinsic and productive, or is it dysfunctional?

An example of a nontraditional application of functionalism is the work of Katz (1960), who developed a functionalist theory of attitudes. According to Katz, an attitude can serve one or more of four functions for the individual: (1) an ego-defensive function (that protects one's ego), (2) a value-expressive function (that allows one to expresses one's values), (3) a knowledge function (that allows one to be informed about matters), and (4) a utilitarian function (that allows one to gain positive consequences or have pleasing experiences). Katz theorized that the strategy that a change agent (e.g., an advertiser, a health educator) should use to change an attitude depends on the function that the attitude is serving. For example, if an attitude largely serves an ego-defensive function, then the advertising or educational strategy used will be different than if that attitude serves a knowledge function. As part of his theory, Katz described the persuasion strategies that are appropriate for each function.

Symbolic Interactionism

Symbolic interactionism is an influential theory in sociology, anthropology, education, and political science, and it incorporates useful perspectives for theory construction. It often is linked to grounded theory frameworks, but its applications extend far beyond such theories. Our discussion of symbolic interactionism follows closely the classic work of Herbert Blumer (1969).

Symbolic interactionism is based on three premises: (1) that people act toward things based on the meanings of those things to them, (2) that meaning is derived from social interactions (i.e., the meanings of objects emerge socially through our interactions with others), and (3) that meaning is the result of an interpretive process used by people to deal with the stimuli that they encounter. The interpretation of stimuli has two phases. First, the actor indicates to him- or herself the things toward which he or she is acting. Second, the actor selects, checks, suspends, regroups, or transforms the meanings in light of the situation. Symbolic interaction occurs when one individual interprets the meaning of another individual's actions or gestures.

According to symbolic interaction theory, interaction involves mutual role taking in which each person tries to see things from the other person's point of view. Only through such role playing can you communicate meanings, because you must understand the frame of reference of the other individual for communication to occur. The self also is a critical concept in theories of symbolic interactionism. People are the object of their own actions, and as a result, they have a self-concept. The self-concept derives from how you think others view you. By having a self, you can interact with yourself and make indications to yourself. This is what sets humans apart from animals.

Human action is the result of a person's interpreting the events that transpire in a situation and then forming plans of action in light of those events. Symbolic interactionism focuses on understanding how individuals construe the environment about them and how they then choose to act upon that environment. The image is not one of an individual being "pushed around" by this cause or that cause, but rather of an active, interpreting individual, choosing to take certain actions.

According to symbolic interactionism, our experience of the external environment is subjective, the result of the meanings we impose on it. Objects in the environment are, in this respect, social creations. They are defined and construed in terms of the images that we have of them in our mind. But symbolic interactionists also recognize that the empirical world can "talk back" and can challenge or resist our images of it. This, in turn, may lead to a new image.

Blumer (1969) argues that social scientists rarely have firsthand knowledge of the empirical worlds they study. Academics who study drug addiction usually have not lived the life of a drug addict nor have they directly experienced the world in a way a drug addict does. Blumer emphasizes the inherent outsider role of social scientists and the need for these scientists to experience as closely as possible the world they are studying. Blumer emphasizes the central role of exploratory analysis, participant observation, and intensive case studies in the pretheory construction stages. Such activities help the researcher form the concepts, images, and ideas used in a theory. Blumer believes it is both arrogant and egocentric of social scientists to impose their view of the world on a phenomenon with which they have limited experience. Social scientists must first gain relevant experiences and get as close to the phenomenon as possible to adequately theorize about it.

As you construct your theories of human behavior, even though you may not adopt a formal symbolic interactionist perspective, it might be useful to think about your theory in these terms. How do the actors in your theory construe their world and the key elements within it? What meanings are they using and how do they communicate these meanings to others with whom they interact? How do they view themselves and how are they forming action plans in light of their interpretations? What are those action plans? How is the setting impacting the meanings that the individuals extract and use? How do the meanings of one actor overlap with, or differ from, the meanings of another actor?

Evolutionary Perspectives

Evolutionary theories of human behavior have a long and somewhat controversial history in the social sciences. Although some might argue that evolutionary perspectives do not constitute a grand theory, they certainly have been used to explain a diverse set of phenomena in many disciplines. The process of evolution can be defined as gradual changes over time in the organic structure of organisms. It was hypothesized to operate by biologists long before Darwin formulated his famous theory of evolution. Darwin's primary contribution was to specify the mechanisms by which evolution occurred. Central to evolutionary approaches are the concepts of *adaptation* and *natural selection*. Adaptation is the product of inherited characteristics of a species that have come into existence through natural selection (Buss, Haselton, Shackelford, Bleske, & Wakefield, 1998). Natural selection has three defining characteristics: (1) variation, (2) inheritance, and (3) selection. *Variation* refers to the fact that organisms within a species vary in a large number of ways. Such variation allows evolution to operate. Some of these variations are genetically transferred from parents to offspring across multiple generations. This is the process of *inheritance*. Some of the inherited attributes are better suited to reproduction and survival than other attributes. Organisms with such adaptive attributes produce more offspring, on average, than those lacking these attributes, because they are more likely to survive and reproduce as a result of them. Across many generations, the result is natural *selection* toward the adaptive attributes. As one example, scientists have speculated that pregnancy sickness (food aversion, nausea, vomiting) during the first trimester is the result of selection to protect the embryo against maternal ingestion of teratogens. In essence, natural selection acts like a sieve that filters out problematic and nonadaptive behaviors (Dawkins, 1996). Over many generations, the result is a set of behaviors that interacts with the environment so as to promote reproduction and survival of species.

In addition to adaptive attributes, evolution produces by-products of adaptation. By-products are attributes that do not solve adaptive problems but which are confounded with attributes that do. For example, the whiteness of bones is a by-product of the fact that bones contain large amounts of calcium, which is associated with bone strength and which has evolved through natural selection mechanisms Finally, there are behaviors that are neither adaptive nor by-products of adaptation, and these are viewed as random noise in the evolutionary process. Evolutionary scientists differ in how they classify behavior into these three categories. For example, some view language as a by-product of large brains, whereas others view language as adaptive in its own right.

Evolutionists emphasize that there are constraints on adaptation so that optimal adaptation seldom results. Adaptation is a slow process that may take thousands of years. Lack of genotypic variation can constrain the course that adaptation takes. Some adaptations have both costs and benefits associated with them, resulting in the evolution of costs as well as benefits to the species. Finally, adaptations do not operate in isolation and must coordinate themselves with other adaptations. This process of coordination can result in compromises that undermine optimal evolution. For all these reasons (time lags, restricted variation, accompanying costs to benefits, and coordination), adaptation does not always function optimally.

Bereczeki (2000) emphasizes three fundamental scientific orientations of evolutionary analysis: (1) adaptation, (2) ultimate causation, and (3) individualism. In terms of adaptation, the evolutionist seeks to identify how a behavior functions historically to increase survival and/or reproduction of the species. The focus is thus on a functional analysis of behavior rather than trying to identify the causes of behavior in a classic cause–effect framework. What functions could the behavior have served in terms of survival and reproduction? In what ways could the behavior enhance survival or increase reproduction? Such functional analysis is a hallmark of evolution theories.

A second important characteristic of evolutionary analysis is a focus on ultimate causation (Bereczeki, 2000). Social scientists explore the causes of behavior at many different levels. Most social scientists are interested in specifying the proximal determinants of behavior, with lesser interest in understanding more distal determinants. By contrast, evolutionists search for the "ultimate causes" of behavior by tracing them to natural selection and the function they serve to promote survival of the fittest.

A third feature of evolutionary analysis is an emphasis on individualism. In traditional theories of evolution, the emphasis was on the idea of group selection; namely, that individuals evolved so as to act for the good of the group. In the 1970s, a group of biologists shifted this focus toward individualism, that is, the view that selection works mainly at the level of the individual so as to enhance an individual's own genetic fitness.

These individual interests, in turn, determine the social institutions that emerge from them.

Evolutionary perspectives on human behavior have critics, and some of the critiques are themselves a rich source of ideas. For example, Lickleiter and Honeycutt (2003) criticized genetic–evolutionary accounts of human development by noting that the cellular machinery of humans probably has more control over genes than genes do over the cellular machinery of humans. Lickleiter and Honeycutt (2003) review evidence for the fundamental importance of developmental regulatory systems and posit an epigenetic approach to behavior that promotes systems-based analysis that goes beyond simple views of causality.

Evolutionary perspectives can be applied to any human phenomena and are not restricted to biologically based behaviors. Evolution, as a metaphor, can be used to construct many interesting questions (e.g., How do relationships evolve? How do attitudes evolve? How do personalities evolve?). How might evolutionary perspectives be used to analyze the growth of a business or an organization? Some businesses and organizations "survive" whereas others do not. Could an evolutionary perspective be adapted to analyze such dynamics?

As an example of the use of the evolutionary framework for the analysis of nonbiological processes, Mahfouz, Philaretou, and Theocharous (2008) analyzed interpersonal attraction using evolutionary metaphors. According to these theorists, variation is represented by exposing oneself to a pool of available mates, selection takes place through the learning of courting mechanisms and socialization processes that maximize the mate's level of physical attractiveness, and retention is reflected in the initiation of strategies for ensuring the successful carrying out of the dating process, with the ultimate purpose being that of generating a long-term commitment. Their analysis yielded several insights into dating that are not apparent in traditional theories of dating. Evolutionary perspectives also have been offered on such diverse phenomena as schizophrenia (Pearlson & Folley, 2008), teachers' negative affect toward gifted students (Geake & Gross, 2008), the development of authoritarianism (Hastings & Shaffer, 2008), violent crime (Barber, 2008), and on-site web-page preferences and navigation (Stenstrom, Stenstrom, Saad, & Cheikhrouhou, 2008), to name a few.

What are the ultimate causes of the phenomenon you are studying? Do any of your variables, concepts, or processes have a basis in the survival of the species? What systems might constrain or regulate the impact of the ultimate causes you identify? What are the roles of time lags, limited variation, accompanying costs, and coordination in explaining your phenomenon?

Postmodernism: A Critical Commentary on Grand Theories

Postmodernism is not a formal system of thought but instead a critical commentary on assumptions frequently made in social science theories and research. It is controversial even among postmodernists. The term *postmodernity* means "after modernity," and reflects a rejection of the basic tenets of an orientation called *modernity*. Modernism is associated with the Renaissance era and evolved essentially at the same time as the capitalist state. As applied to social science, Gergen (2001) argues that there are three major tenets of modernism that postmodernists question. First, in traditional modernist thought, humans are seen as having the capacity for reasoned and rational deliberations. According to this viewpoint, the thoughts and subjective knowledge of people are key to understanding behavior because a person's thoughts mirror reality, albeit sometimes imperfectly so. Postmodernists question the concept of rationality as traditionally conceived. According to postmodernists, a person is judged to be rational by a community only if he or she adopts the particular codes of discourse common to that community. For example, scientists are judged as being rational by the scientific community only to the extent that they adopt the discourse rules of the community of science. From this point of view, rationality is relational rather than absolute and can shift from one discourse community to another. Being rational is nothing more than a form of communal participation. Rationality is relative.

Second, in the modernist tradition, distinctions are made between the inner world of the mind and the external world of the material. An objectively knowable and rational world is believed to exist that contains systematic cause-and-effect relationships. In contrast, postmodernism emphasizes a social constructionist viewpoint (see Chapter 2). To speak of the world at all requires language. The words of a language are not mirrors of the world but rather arbitrary parts of a language system. To describe an external reality of causal relations reflects nothing more than participation in a language that draws upon the repository of thoughts and knowledge of a particular cultural tradition. Reality, like rationality, is relative.

Third, modernism holds that language is a means of conveying knowledge to others and hence, it is "the bearer of truth" (Gergen, 2001). The postmodernist view is that language is the result of a cultural process that is generated within the context of human relationships. Language becomes meaningful not from its subjective and objective underpinnings but rather from its use in action.

Postmodernists are suspicious of authoritative declarations or statements of "truth." They view such declarations as oppressive and silencing, especially for oppressed groups, such as females, members of certain ethnicities, and third-world peoples. For postmodernists, understanding is individualized, and there are an infinite number of possible interpretations. To declare one interpretation as correct is to oppress others. Gergen (2001, p. 808) captures this philosophy as applied to psychology as follows:

> As psychological theories are exported to the culture more generally, what are the reverberations in cultural life? When one holds that the primary ingredients of the mind are cognitive, when one views behavior as genetically prepared, when one distinguishes between pathology and normalcy, which doors are opened within the culture, and which are shut? . . . Who is helped and who is hurt when psychologists distinguish between the intelligent and unintelligent, the pathological and normal, the prejudiced and unbiased? What form of

> culture is created when exploitation, infidelity, and rape are viewed as biologically prepared actions of the male?

Gergen argues that it is important for theorists to think about how they are framing problems and how this might impact other members of society. This, in turn, might impact the choices one makes in framing a problem. Gergen goes on to emphasize the liberating role of postmodernism for psychological theory:

> If scientific descriptions and explanations of the world are not demanded by the nature of the world itself, then one is released from the shackles of the taken for granted. Most importantly, one is invited into a posture of theoretical creativity. Scientists are liberated from the task of being mere mirror holders to the world as it is and challenged to articulate new and potentially transformative conceptions. The task is not simply that of describing what currently exists but of creating intelligibilities that may foster worlds to come. Metaphorically, the function shifts from that of scribe to poet. (p. 810)

Postmodernists vary in the extremity of their positions within postmodern philosophy. Some reject any attempt at theory construction as futile because everything is relative, whereas others view theory construction as another form of dialogue among an infinite number of possible dialogues. The postmodernist approach emphasizes the importance of understanding the cultural context in which behavior occurs and the cultural context in which interpretations of behavior are made.

Postmodernists often engage in a process called *deconstruction*, a strategy of analysis associated with the philosopher Jacques Derrida (Thomassen, 2006). Derrida objects to attempts to turn deconstruction into a coherent analytic strategy, but there are themes associated with it that many have sought to depict. As applied to textual narratives, deconstruction focuses on identifying what the text represses—what it does not say—and its incongruities. Rosenau (1992) suggests that deconstruction involves (1) finding an exception to a generalization and pushing it to the limit so that the generalization is undermined, (2) interpreting arguments in a text in their most extreme form, (3) avoiding all absolute statements and seeking to make statements that are striking, (4) denying the legitimacy of categories because one can always find exceptions, (5) accepting nothing and rejecting nothing, and (6) using new and unusual terminology to avoid the surplus meaning associated with more established terms.

Although we have difficulty with several tenets of postmodernism, we find it healthy at times to think about concepts, relationships between concepts, and processes from this perspective. As you choose variables, concepts, and processes to study and elaborate theoretical narratives about, what groups or viewpoints are you affecting by the way you frame the underlying issues? How are your viewpoints filtering your questions and observations? If you rejected the three major tenets of modernism (that people are rational, that there is an objective reality, and that language is a mirror of that reality) and adopted the perspective of the postmodernist, how would this perspective shape the kinds of variables, concepts, and processes you study? How would this impact your explanations of events? What would happen if you tried to deconstruct the narrative account of your theory? Sometimes we write out a theory and then subject it to the deconstruction process, with the result being new insights or perspectives.

FRAMEWORKS USING METAPHORS

In this section we describe two general systems of thought that make use of biological metaphors. The first is the framework of neural networks, which combines causal analysis and process analysis in ways that are distinct from the theory construction processes described in previous chapters. The second framework, systems theory, has its roots in metaphors from physiology. Physiologists view the human body as a collection of interacting systems, each with its own combination of functions and purposes. These include the nervous system, the musculoskeletal system, the circulatory system, the respiratory system, the gastrointestinal system, the integumentary system, the urinary system, the reproductive system, the immune system, and the endocrine system. These systems coordinate to promote the survival and effective functioning of individuals over time. In systems theory frameworks, one analyzes a phenomenon in terms of the system in which it is embedded and then how that system interacts with other systems to impact outcomes of interest.

Neural Networks

Neural network theories do not have the “track record” of the grand theories described in the previous section, but they are becoming increasingly influential in the social sciences. They use neural mechanisms as metaphors for analyzing a range of behaviors and mental phenomena. We present the core elements of a neural network approach in abstract form, weaving in an example to make it more concrete. Neural networks are popular in the analysis of learning, memory, and information processing. They can be readily applied to decision making, attitude formation, prejudice, and most any phenomenon that involves the processing of information by individuals. The framework also can be adapted to describe relationships between organizations and organizational structures.

The basic unit in a neural network theory, a *neuron*, is connected to other neurons. The neurons are organized into layers, with the first layer representing the *input layer* and the last layer representing the *output layer*. The layers of neurons in between are called *hidden layers*. For example, in a neural network model of impression formation, each neuron at the input level might be a unit reflecting the different pieces of information that a person is given about a stranger. Suppose we describe a teacher to a prospective student using three personality traits and then ask the person how much he or she would like the teacher. Each piece of trait information is fed to a separate neuron in the first layer and then these propagate to the neurons in the second layer, and so on, until the final output layer is reached. Essentially, the neurons at the first level are predictors and the neurons or units in the last layer are outcomes. The input and output layers are usually observable (the input information and the output response) and the layers in between usually are not observable; hence the terminology *hidden layers*.

Any given neuron can be active or inactive. Activation is binary in character; that is, it takes on the value of 1 (active) or 0 (not active). The presentation of the trait information activates some neurons, called input neurons, but not others. An activated input neuron reaches a neuron at the next layer through threshold units. There is a separate threshold unit for each output neuron. A threshold unit can be active or inactive. When the threshold unit becomes active, it "fires" like a synapse to activate the next neuron it is connected to. The threshold unit has an activation index that reflects its accumulated energy toward activation. The threshold unit also has a threshold value relative to its activation index, such that when the threshold value is exceeded by the activation index, the threshold unit fires, thereby activating the next neuron. Multiple neurons can feed into the threshold unit, each through a weighted connection, to impact its activation index. The connection can be either excitatory (i.e., the input neuron can either increase

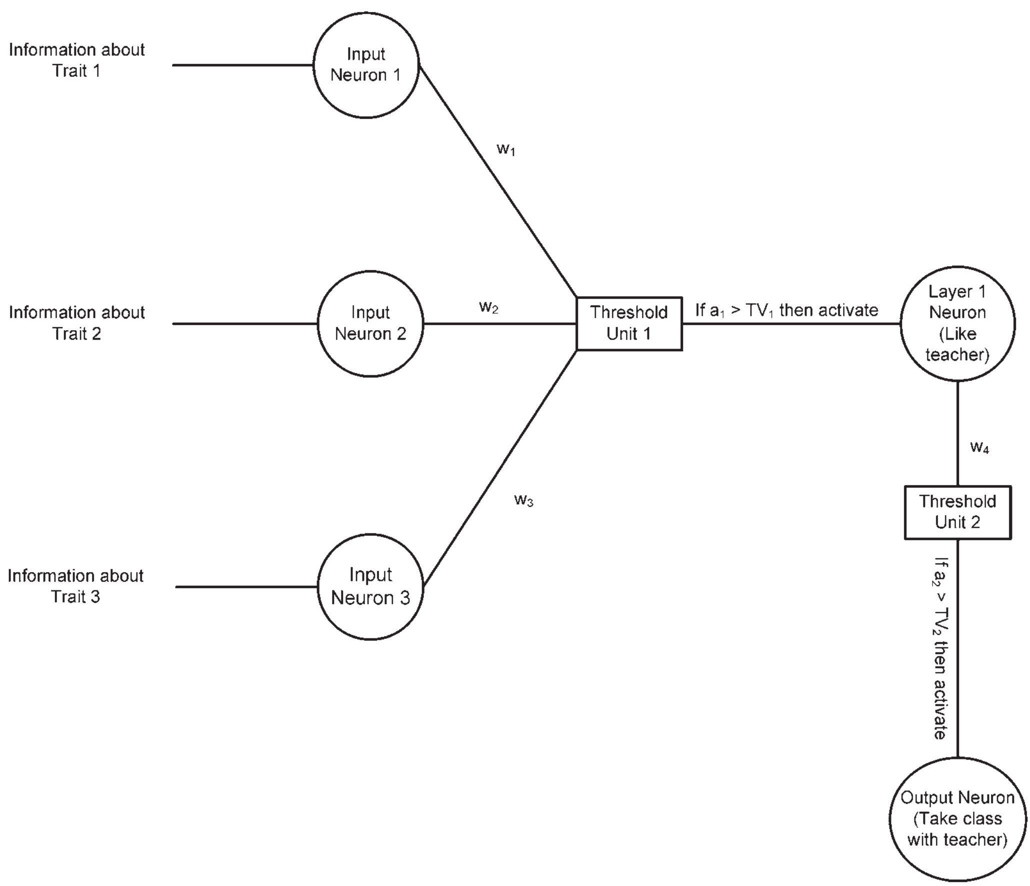


FIGURE 11.1. A neural network.

the activation index) or inhibitory (i.e., the input neuron can decrease the activation index). The larger the weight of the connection, the greater the contribution of the neuron to the threshold unit.

Figure 11.1 presents a simple schematic of the process. Information is presented about traits 1, 2, and 3, and these activate input neurons 1, 2, and 3, respectively. These neurons are linked to a hidden layer neuron (liking for the teacher), and this neuron is linked to an output neuron (taking a class from the teacher). Input neurons 1, 2, and 3 feed into the threshold unit for the layer 1 neuron. Their contribution to the threshold unit is reflected by the weights of the connections (w_1, w_2, and w_3). The more the weight for a given neuron deviates from 0, the greater the contribution it has for potentially activating the neuron to which it is connected. For example, the neuron for trait information 1 might have a larger weight (w_1) than the neuron for trait information 2 (w_2). The activation index of the threshold unit is signified by a_1 and its value is determined by the weighted combination of the connections from each of the input neurons. If the value of a_1 exceeds the threshold value of the threshold unit (TV_1), then a synaptic-like firing occurs, and the layer 1 neuron is activated. The same process occurs linking the layer 1 neuron to the output neuron.

This is a simplistic representation that is intended to convey the basic processes involved in a neural network. The example we described could be expanded so that other layer 1 neurons impact the output. Also, through the acquisition of experience, the weights of the connections (w_1, w_2, and w_3) can change. Thus, neural networks can "learn" as the strength of connections increases or decreases. For example, if a person is told by several different individuals that a teacher has a particular trait, then this repetition may strengthen the connection between that trait's input neuron and neurons in other layers. Neural network theorists have described a wide variety of learning strategies that neural networks can incorporate. A specialized branch of mathematics and statistics has evolved for the analysis and structuring of neural network approaches. Neural network models take many forms and can be quite complex in nature.

Interesting examples of neural network models abound in the social sciences. For example, Marshall and English (2000) applied neural network analysis to model the association between social workers' overall assessments of the risk status of children in child protective services and information about 37 separate risk factors from the state of Washington's risk assessment matrix. Marshall and English found complex relationships between the risk information available and caseworker judgments of risk that were not evident from more traditional modeling strategies. The neural network analysis indicated that only a small number of the 37 risk factors were actually used by caseworkers when making their judgments and, as a result, the state of Washington revised its risk assessment matrix. In addition, new training methods were implemented to reemphasize the importance of certain risk factors (e.g., history of domestic violence) known to be related to recidivism but which the neural network analysis showed were not being used by caseworkers in their decision to place a child in foster care. Neural networks map closely onto another type of model, called connectionist models.

You may find it constructive to reconceptualize your theory using a neural network metaphor. What would constitute the neurons for the input layer in your theory? What would be the neurons in the output layers? What are the hidden layers and how are they organized? What neurons are connected directly to other neurons? How much does it take to activate a neuron (as reflected by the threshold value of the threshold unit connected to it)? Do some of the threshold units in your network have higher threshold values than others, and if so, why? Do some neurons contribute more to a threshold unit than other neurons (as reflected by their different weights, w_i)? What can you conjecture about these differential weights and the sources of the differences? There is an extensive literature on neural network modeling that can assist you in answering these and other questions (see Abdi, Valentin, & Edelman, 1999, for an introductory treatment)

Systems Theory

General systems theory was introduced by Ludwig von Bertalanffy in the 1940s, who presented the theory using biological metaphors. Systems thinking is pervasive in everyday life, as evidenced by such concepts as the health care system, the family system, political systems, ecological systems, and banking systems. General systems theory has been subject to a wide range of interpretations. Sadovski (1974) found over 35 different definitions of what a "system" is in scientific writings. It is impossible to convey a unified view of systems theory in light of this. Indeed, one of the main complaints against general systems theory is that it is too vague. We emphasize here systems concepts that reflect different ways of thinking about phenomena relative to the other frameworks discussed in this chapter.

At the simplest level, one can think of a system as an organized entity whose interrelated elements interact with one another so as to achieve some common goal. There are many types of systems: A *static* system remains invariant over time; a *dynamic* system is constantly changing; a *closed* system has limited or no interaction with the environment or other systems; an *open* system interacts with the environment or other systems, exchanging inputs and outputs.

Systems theory adopts a holistic approach rather than a reductionist approach to analysis. It seeks to gain insights into the functioning of the overall system by examining the interrelationships and connections between the component parts of the system. It is not enough, for example, in a family system to separately analyze the mother, the father, the son, and the daughter. Rather, one must examine how these elements interact and the interconnections between them to more fully understand family functioning. Systems theory emphasizes understanding processes rather than causes. Instead of analyzing *why* families communicate, for example, the focus is on *how* families communicate and how this changes as a function of settings and contexts.

Systems theory also emphasizes the concept of feedback. *Feedback* refers to a circular process in which input is transformed by the system into output, and the output is brought back to the system as input. Feedback loops allow a system to regulate its behavior and are integral to system analyses.

Many systems have regulatory mechanisms that strive to maintain equilibrium. An analogy is a thermostat that heats a home during winter. The equilibrium point is the temperature at which the thermostat is set. The thermostat monitors the room temperature and sends a signal to the heater to turn on if the room's temperature falls below the equilibrium point and for the heater to turn off if the temperature is above the equilibrium point. Systems analysis examines such regulatory mechanisms, emphasizing the underlying processes that are taking place, the feedback loops that are operating within the system, and the continual adjustments that are being made among the many system elements. Another concept emphasized in systems theory is that of *equifinality*, which refers to the idea that there may be several equally effective ways to achieve a goal (Van de Ven & Poole, 1995).

Brinberg and McGrath (1985) discuss three facets of systems that theorists can consider when thinking in terms of systems theory, either for the system as a whole or at various levels of a system. First is *system well-being*, which refers to the identification of conditions and behaviors that threaten the health and safety of the system or that constrain the positive development of the system. Second is *system task performance effectiveness*, which refers to conditions and barriers that hinder the system from carrying out its tasks or achieving its goals. Third is *system cost*, which refers to expenditures of energy and resources in pursuit of task performance.

There are many examples of systems theory approaches in the social sciences. In the field of criminal justice, for example, Van Gigch (1978) used systems theory to analyze major components of criminal justice (police, courts, and corrections) and the agencies and institutions that serve those components. Van Gigch identified the inputs and outputs of each institution and argued that these were regulated by the needs of the overall criminal justice system. His theory emphasized that the criminal justice system as a whole was greater than the sum of its parts. He also argued that the criminal justice system is, itself, a subsystem within larger political, economic, educational, and technical systems, all of which impact how the criminal justice system functions. Kraska (2004) argued that the systems approach to criminal justice facilitates macro-level analyses of criminal justice and offered numerous suggestions for reform based on such analyses. Bernard, Paoline, and Pare (2005) offer a theory of criminal justice based on nine propositions that derive from systems theory perspectives.

As another example, in the 1980s, questions were raised in Canada about the health risks of a herbicide called Alachlor. Three different stakeholders performed what each characterized as objective, scientific assessments of the risks of Alachlor, but each came to a different conclusion. Hatfield and Hipel (2002) used systems theory to isolate the bases of the differing conclusions and advocated for the use of general systems theory in risk assessments more generally.

Systems theory is widely used in analyses of families (Freeman, 1993; Wood, 2002) and organizations (Takahashi, Kijima, & Sato, 2004). Thinking about your theory, what happens if you think about matters in terms of processes performed by systems rather than variables? What insights do you gain by describing how things function rather than trying to describe why things function as they do? Are there feedback loops? Are their regulatory mechanisms? In what ways is the whole larger than the functioning of the individual parts? As noted earlier, the human body has a nervous system, a skeletal system, a circulatory system, a respiratory system, and a multitude of other systems. These systems are interrelated and work together toward the common goal of keeping a person alive. Can you use this as a metaphor for the phenomena that you are studying? What are the core systems in your theory? How are they interrelated and how do they interact with one another? How does one system depend on the other? What are the functions of each system and what processes take place in the system to allow these functions to be fulfilled?

FRAMEWORKS EMPHASIZING STABILITY AND CHANGE

Although all of the frameworks discussed can be applied to the analysis of change, there are perspectives on stability and change worth mentioning that stand apart from these systems of thought. Van de Ven and Poole (1995, 2005) describe four conceptual orientations to process-based analyses of change. First, there is a *life-cycle model* in which the developing entity (an individual, a couple, a social group, an organization) is thought to move through a set of preconfigured or preprogrammed phases from beginning to end. For example, human development is often thought to progress from infancy to childhood to adolescence to young adulthood to adulthood to older adulthood, with noteworthy cognitive, moral, emotional, social, and physical changes associated with each phase of development. A second framework for process analyses of change identified by Van de Ven and Poole is a *teleological model* that conceptualizes individuals as moving toward a goal or end state in a purposeful way; the person specifies a goal and takes action to reach it. In a third framework, a *dialectical model*, the person or entity is viewed as being influenced by opposing events and forces, some promoting change and others promoting stability. Change is interpreted in the context of the balance of power between these opposing forces. The final framework, an *evolutionary model*, emphasizes cumulative change, as discussed earlier in the context of evolutionary perspectives on theory construction.

These four approaches to process-based analyses of change are not mutually exclusive. All four mechanisms, or some combination of them, can operate to produce or characterize change. Van de Ven and Poole (1995) described 15 different frameworks that involve various combinations of the four mechanisms.

Related to the life-cycle model of Van de Ven and Poole (1995) are stage theories of change. These theories conceptualize people as moving through sequential stages in their progress toward some outcome or end state. Unlike the life-cycle model, however, progression through the stages is not necessarily predetermined. A stage theory specifies what the stages are, the criteria used to classify a person (or other entity) as in one stage as opposed to another stage, and the processes or requirements that must be mastered or achieved to move from one stage to the next. Stage theories are common in such disciplines as developmental science, health, and organizational studies. As an example, in the area of smoking cessation, the transtheoretical model identifies five stages through which a person is presumed to progress (Prochaska, DiClemente,

& Norcross, 1992). The first stage is the *precontemplation stage*, where a smoker has no intention of quitting smoking in the next 6 months. In the *contemplation stage* a smoker is thinking about quitting some time in the next 6 months but does not plan to quit in the next month. The *preparation stage* marks the smoker's intention to quit within the next month and includes at least one unsuccessful 24-hour quit attempt in the past year. In the *action stage* the smoker successfully quits smoking for any time between 1 day and 6 months. After 6 months of being smoke free, the ex-smoker is said to have reached the *maintenance stage*. According to the transtheoretical model, the type of intervention that one uses to help a person stop smoking differs depending on the "stage of change" in which he or she is. For example, someone in the precontemplative stage will need different information and support than someone in the maintenance stage.

Stage theories differ in at least five ways. First, some stage theories permit the entity (e.g., a person, a family, an organization) to revert back to previous stages, whereas other stage theories dictate a forward progression only. Second, some stage theories require that entities spend a fixed minimum amount of time in a stage, whereas other theories permit movement through stages with no time constraints. Third, some stage theories define stages in terms of categories along a single continuum, whereas other stage theories define the stages multivariately. For example, the first three stages of the transtheoretical model represent a trichotomization of the intention to quit smoking along a time dimension (never, within 6 months, within 1 month). By contrast, Freud's psychoanalytic theory of the stages of psychosexual development (the oral stage, the anal stage, the phallic stage, the latency stage, and the genital stage) involve distinctly different dimensions at each stage. Fourth, some stage theories permit entities to skip an intermediate stage, whereas other theories require each entity to progress through each stage in sequence. Finally, some stage theories characterize movement from one stage to another in terms of large, dramatic shifts in behavior, whereas others view the passing from one stage to the next as a more gradual process.

Dixon and Moore (2000) discuss stage theories from the perspective of the developmental ordering of two or more skills or types of knowledge. Consider two different but interdependent skills, *A* and *B*. Dixon and Moore argue that if acquisition of one skill is developmentally dependent on the acquisition of another skill, then this, for all intents and purposes, is a stage theory. If the development of skill *B* is dependent on the development of skill *A*, then skill *A* is the first stage and skill *B* is the second stage. Dixon and Moore suggest different models of developmental ordering for stage theories. *Developmental synchrony* occurs when both skills begin development at the same time and basically develop at the same rate. One skill is dependent on the other, but the dependence is instantaneous. *Partial developmental priority* occurs when the two skills start at the same time, but one develops at a faster rate than the other. The level of skill *B* might be dependent on the level of skill *A*, but there is a lag between the acquisition of a level of skill *B* and the acquisition of a certain level of skill *A*. The lag may not be complete in that only a certain level of mastery of skill *A* is required for skill *B* to start developing. *Complete developmental priority* occurs when skill *A* must be fully mastered before skill *B* can begin to develop.

There are many examples of stage theories in the social sciences, with the transtheoretical model described above being one. As another example, Duck (1982) developed a stage-like theory of relationship dissolution that consists of six phases. First, there is a *breakdown phase* where one or both partners become distressed over the way the relationship is being conducted. This eventually leads to an *intrapsychic phase* that is characterized by brooding about the relationship. Nothing is said to the partner at this time, but the focus is on one's feelings that the relationship is damaged. This phase, which is characterized by uncertainty, anxiety, and hostility, is followed by the *dyadic phase*, where partners confront each other and talk through their feelings about the relationship. If a decision is made to break up, they move to a *social phase*, where they tell others of their decision and seek social support. This, in turn, leads to the *grave-dressing phase*, where the relationship is seen as being "dead," with the recognition that it must be "put away." During this phase, the individual builds a narrative about what the relationship was, reinterpreting it and labeling it in light of his or her current feelings. Finally comes the *resurrection process* in which the individual re-creates a sense of his or her social value and defines what he or she wants out of future relationships. This theory was recently updated and enhanced by Rollie and Duck (2006), who emphasize predominant processes throughout the dissolution experience and elaborate on the role of communication at each phase.

In your theory construction efforts, consider if you want to conceptualize any of your variables or phenomena using a life-cycle model, a teleological model, a dialectical model, or an evolutionary model. Consider imposing a "stage" framework. If you think in these terms, elucidate the criteria for defining stages and make decisions about the evolution of stages using the dimensions we discussed on how stage theories differ. Can people revert back to previous stages? Must people spend a fixed minimum amount of time in a stage? Can your stages be defined in terms of a single dimension or must they be defined multivariately? Can people skip a stage? Are the shifts from one stage to the next gradual or dramatic? What is the developmental ordering of the "skills," "knowledge," or "events" in your theory in the spirit of the Dixon and Moore analysis?

PSYCHOLOGICAL FRAMEWORKS

Several influential frameworks in psychology offer unique perspectives relative to the frameworks we have discussed thus far. We focus on two systems of thought, reinforcement theory and positive psychology. Although there are other theories that are either more contemporary or more influential historically, we focus on these two frameworks because they complement well those that have already been discussed.

Reinforcement Theories

Reinforcement theory achieved prominence primarily in psychology, but it also has been applied in education, sociology, political science, and public health. Reinforcement theories emphasize the concepts of *stimulus* and *response*. A stimulus is an external or internal event that leads to the performance of a behavior, which is termed the response. Stimuli have three functions: (1) elicitation, (2) discrimination, and (3) reinforcement. An *eliciting stimulus* evokes an instinctual, natural response, such as the sight of food stimulating salivation. A *discriminative stimulus* does not directly elicit a response but sets the stage for the response. It signals to people that they should respond in a certain way. For example, when a child is eating at the dinner table, he or she may exhibit different table manners when the mother is present than when the mother is absent. The mother, in this case, is a discriminative stimulus. *Reinforcing stimuli* occur as a positive or negative consequence of a response. Reinforcement theories focus on identifying the roles of relevant stimuli as being eliciting, discriminatory, and/or reinforcing. For example, if a child acts aggressively, reinforcement theory focuses on identifying what stimuli in the environment are serving to positively reinforce such behavior and what discriminative cues are present that signal the permissibleness of aggressive behavior.

In some (but not all) reinforcement theories, the concept of *drive* is important. Clark Hull (1943) proposed that there is an underlying, unlearned, biological source of core needs related to metabolic processes, such as the need for food, water, and sexual contact. When a person is deprived of the satisfaction of these needs, a drive force is created that activates behavior in the direction of trying to satisfy them. When satisfaction of the need is obtained, reduction in the drive occurs. Dollard and Miller (1950) identified two types of drives: (1) primary or innate drives and (2) secondary drives that are the product of learning (e.g., a drive toward monetary rewards or a drive toward verbal rewards). Another facet of reinforcement theory is to analyze the fundamental drive states, both primary and secondary, that may underlie behavior.

A fundamental tenet of reinforcement theories derives from Thurstone's *law of effect*. This law states that if positive consequences follow an individual's response to a stimulus, then he or she will probably repeat that response in the presence of the stimulus on future occasions. Conversely, if a response to a stimulus is followed by negative consequences, then the individual will avoid repeating that response to the stimuli in the future. One can increase or decrease the probability of a behavior occurring by manipulating reinforcers of the behavior. *Positive reinforcers* are stimuli that serve to strengthen the response; *negative reinforcers* are stimuli that strengthen the response when they are removed. For example, if an aversive stimulus is removed in response to a behavior, then the removal of that stimulus serves as a negative reinforcer. *Punishments* are stimuli that reduce the strength of the response. When analyzing behavior, reinforcement theorists also analyze what stimuli may be operating as reinforcers for the behaviors and what stimuli might be operating to punish the behavior. By identifying these stimuli, one begins to understand the bases of the behavior.

The effectiveness of a reinforcer in shaping behavior depends, in part, on the schedule of its administration. There are two broad types of reinforcement schedules: continuous and intermittent. Continuous reinforcement occurs when a behavior is reinforced each time it occurs. Research suggests that continuous reinforcement is the fastest way to establish new behaviors or eliminate undesired behaviors. Intermittent schedules occur when only some instances of a desired behavior are reinforced.

Stimulus generalization is an important process in reinforcement theories. It refers to the situation in which a novel stimulus evokes a response that was previously learned in relation to a different though similar stimulus. For example, if two situations are very similar to each other and individuals learn to respond in a certain way in one of the situations, they probably will respond the same way in the other situation. *Discrimination* refers to an individual's ability to differentiate stimuli and respond differently to them. Generalization and discrimination go hand in hand: They allow individuals to economize their behavior by generalizing stimuli, on the one hand, yet make important distinctions and respond differentially, on the other hand. Another key point of analysis for reinforcement theories is the analysis of stimulus generalization and how reinforcement histories from one context may generalize to another context.

Extinction is another important process in reinforcement theories. It refers to the progressive decrement in a response under continual nonreinforcement. Thus, if a response is no longer reinforced in the presence of a stimulus, then the response will eventually extinguish. Some people extinguish responses quicker than others. Resistance to extinction is a function of (1) the amount of prior reinforcement, (2) the strength of the drive to obtain the reinforcer during extinction, (3) the amount of work or effort involved in performing the response, and (4) the schedule that was used to reinforce the behavior initially.

Although reinforcement theories have their critics, we often find it useful to "put on the hat of a reinforcement theorist" and think about how such a scientist would analyze the outcome variables in which we are interested. More often than not, explanatory constructs suggest themselves that otherwise would not have. What stimuli are eliciting the behavior? What reinforcers are operating to encourage the behavior? Are there any potential negative reinforcers? What punishments are operating to discourage the behavior? Are there any basic drive states that may be impacting the behavior? What are the operative reinforcement schedules—that is, what is the dynamic process by which reinforcers are being administered? Are there discriminative stimuli operating that point to one type of behavior in the presence of those stimuli but a different behavior in the absence of those stimuli? Is the behavior impacted by a person's history in similar situations to the ones we are studying? To what situations might the effects of variables generalize? If we want to extinguish a behavior, how would we go about removing reinforcers? What would make some people more (or less) resistant to extinction? Trying to answer these questions, even if you are not a reinforcement theorist, can often be a rich source of ideas and perspectives.

Reinforcement theory and the many variants of it are widely used in behavioral interventions to impact inappropriate and destructive behaviors in children, adolescents, and adults (e.g., Durand, 2002; Kazdin, 2008; Patterson, 1975). These interventions identify the behaviors that need changing and then analyze these behaviors in terms of the questions noted above. As relevant reinforcers affecting the behavior are identified, they are either removed or new reinforcers are added. Or, if a desired reinforcer is being administered only intermittently, then a more effective reinforcement schedule is implemented. Or, if the child, adolescent, or adult is responding inappropriately only in the presence of a certain discriminative cue, then attempts will be made to eliminate that cue. If the child, adolescent, or adult performs appropriate behaviors in one setting, then methods for generalizing that response to other situations will be introduced.

Reinforcement theory is no longer a cornerstone of mainstream psychology, which, ironically, makes it all that more useful for generating unique perspectives on behavior relative to more dominant theories. Reinforcement theories typically are used to analyze the behavior of individuals, but there is nothing to prevent you from thinking about groups, organizations, or other larger units in these terms. For example, when studying relationships within or between hierarchically structured organizations, how does one department or unit affect another from the perspective of reinforcement theories? Not all of the concepts we outlined apply, but many of them do. Again, try to use the different theoretical perspectives we are presenting in creative ways that go beyond the routine application of them.

Humanism and Positive Psychology

A longstanding approach to the analysis of human behavior is that of humanism, which recently has been reinvigorated by a movement in psychology called "positive psychology" (Bacon, 2005; Kimble, 1984). Positive psychology focuses on understanding strengths and virtues, positive emotions, and positive institutions within society (Seligman & Csikszentmihalyi, 2000). It seeks to understand what factors contribute to these phenomena and how they can be nurtured and developed. It has its roots in psychology in the classic work on client-centered therapy as espoused by Carl Rogers (1961). This theory emphasizes people's innate tendencies to seek self-actualization and health, when conditions permit.

Seligman (2002) presents a positive psychology framework that emphasizes the concept of *strengths*, which he believes are distinct from other psychological constructs, such as traits and abilities. According to Seligman, strengths are valued in their own right and are characteristics or attributes that parents want for their children. Strengths tend to be valued in all cultures. Examples include creativity, wisdom, bravery, leadership, humility, and integrity. Peterson and Seligman (2004) present a taxonomy of 24 strengths that are organized in terms of six higher-order virtues. Bacon (2005) presents an alternative taxonomy that emphasizes focus-oriented strengths and balanced-focused strengths. *Focus-oriented strengths* emphasize individual growth and development, whereas *balance-oriented strengths* emphasize bringing about harmony within the self and between the self and others.

A major contribution of positive psychology is that it has focused theorizing on outcome variables that have a positive character to them. Whereas much of social science is aimed at understanding negative outcomes, such as unhealthy behavior, prejudice, family dissolution, suicide, drug use, HIV-related behaviors, conflict, and survival under adversity, positive psychology shifts the focus to how normal people flourish under normal or benign conditions. Thus, this framework impacts the kinds of outcome variables on which you might choose to focus your theories.

Positive psychology also has implications for the kinds of causes you might focus on and the strategies you use to develop interventions. This point is illustrated in research on adolescent problem behaviors. Most interventions aimed at addressing such behavior are problem-specific. For example, there are school-based interventions aimed at preventing sexual risk taking, smoking, drug use, and alcohol use. These programs focus on the beliefs and attitudes of adolescents toward the problem behavior in question, peer influences on these behaviors, and opportunity structures that make it easier to perform the behaviors. An alternative approach to intervention design is one that emphasizes not the problem behaviors of youths but rather the positive features of adolescent development (Catalano, Berglund, Ryan, Lonczak, & Hawkins, 2002). Known as positive youth development (PYD) programs, these interventions assume that the same individual, family, school, and community factors that influence positive outcomes in youth also impact youth-related problem behaviors. PYD interventions focus not on problem behaviors but instead on promoting bonding, resilience, spirituality, social skills, moral competence, self-efficacy, belief in the future, prosocial norms, and other general orientations toward life that serve the positive development of youths. The idea is that youths who have such positive orientations will be less likely to engage in problem behaviors. In this case, a "positive" focus on human behavior impacts the kinds of explanatory variables one chooses to emphasize relative to the more traditional problem-related beliefs, attitudes, and norms.

We believe that positive psychology perspectives are worth considering when thinking about outcome variables and factors that impact them. If your theory is focused on a negative outcome state, might you also consider adding positive states to your theory? When thinking about ways of reducing a negative state, what positive factors might you try to enhance to do so? Instead of reducing the negatives, might your theory be expanded to increase the positives?

FRAMEWORKS INSPIRED BY METHODOLOGY

Multilevel Modeling

Social scientists in many disciplines have long advocated the importance of incorporating into theories the broader contexts in which individuals behave. Research adopting such perspectives has increased exponentially with the recent advent of a statistical method of analysis called multilevel modeling. To be sure, the rudiments of multilevel modeling have been around for decades, but it is only in the last 15 years or so that reasonably user-friendly software has become available to make sophisticated multilevel modeling accessible to social scientists. This, in turn, has shaped the kinds of multilevel questions that scientists ask. In this section we describe multilevel thinking with the idea that bringing such frameworks to bear on the phenomena you study may provide you with new insights and perspectives.

Consider the case in which we wish to explain student performance on math achievement tests. One set of explanatory variables focuses on individual differences, such as gender. A theorist might offer the proposition that boys score, on average, higher than girls because of a host of socialization factors, which we do not elaborate here in the interest of space. This proposition can be diagrammed using a path model (see Chapter 7), as in Figure 11.2a.

Students attend different types of schools, and it is possible that the type of school they attend also impacts their test performance. For example, students who attend private schools might perform better on math achievement tests than students who attend public schools by virtue of the smaller class sizes and more individualized attention they receive. The type of school is a contextual variable in the sense that students are "nested within" different types of schools. This influence can be diagrammed in conjunction with the gender effect, as in Figure 11.2b.

Finally, the theorist might argue that the type of school moderates the impact of gender on math achievement scores. Specifically, the theorist reasons that in private schools, where students receive more individualized attention, the gender difference between boys and girls is lessened when compared with public schools. This yields the revised path diagram in Figure 11.2c.

Figure 11.2c (and Figure 11.2b) represents a two-level model. Units in the first level (students) are nested within units of the second level (schools), and the characteristics of the second-level units (i.e., characteristics of the schools) are thought to influence the

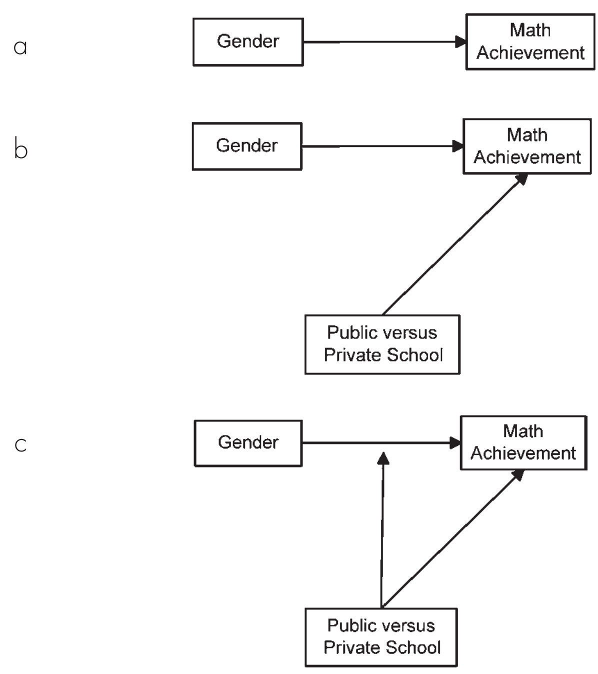


FIGURE 11.2. Examples for multilevel modeling. (a) Level 1 analysis; (b) multilevel model; (c) multilevel model with moderation.

level 1 outcome variable (math achievement) and/or influence the way in which level 1 characteristics (gender) impact the outcome variable.

Multilevel modeling invokes a frame of mind in which theorists think about explanatory variables at different levels, focusing on cases where level 1 units are nested within level 2 units. There are many such examples. Students are nested within schools, and characteristics of both the students and the schools can influence performance on tests. Employees are nested within organizations and characteristics of both the employees and the organization can influence employee behavior. Patients are nested within hospitals and characteristics of both the patients and the hospitals can influence patient recovery. Infants are nested within families and characteristics of both the infants and the families can influence infant development.

The above are examples of two-level models. Multilevel theorists also work with three-level models in which level 1 units are nested within level 2 units, which, in turn, are nested within level 3 units. For example, students are nested within schools, which, in turn, are nested within counties. Characteristics of the students, the schools, and the counties all can influence performance on tests. As another example, in organizational analyses, employees might be nested within departments, which, in turn, are nested within organizations. Characteristics of the employees, the departments in which they serve, and the organizations all can influence job performance.

The relationships between variables at the different levels can be expressed in a variety of ways, such as through the use of propositions in narratives, path diagrams (per Figure 11.2), and/or via mathematical equations. The essence of multilevel modeling, however, is a focus on nested units and the explanatory variables at each level of nesting.

In some theories, level 2 variables are treated as aggregations of level 1 variables. For example, the morale of a family (level 2) might be the average morale across the different family members (level 1), and the cohesiveness of a family (level 2) might be the variance of the morale across family members (e.g., in some families, all members have the same level of morale, whereas in other families, morale differs considerably from one family member to the next). In other cases, the level 2 variables are defined independently of the level 1 variables, such as the case where a school is public or private, or whether the school has a strict or lax policy for expelling students caught with marijuana.

There is a robust literature on multilevel analysis. As you think about the phenomena in which you are interested, consider if they are amenable to multilevel analysis. What are your level 1 units, what are your level 2 units, and what are your level 3 units? What explanatory variables are relevant at each level? Within a given level, what are the causal relations between the variables? What are the causal relations across levels? If you adopt a causal framework, are there moderated effects? Are there mediated effects? Are your level 2 variables simple aggregates of your level 1 variables, or are they defined independent of them? Are your level 3 variables simple aggregates of your level 2 variables, or are they defined independently of them? The multilevel framework encourages you to think about explanatory variables in ways that differ from more traditional forms of analysis and that may therefore enrich your theory.

A unique application of multilevel modeling is the analysis of "growth curves." This application reconceptualizes temporal dynamics by viewing individuals as level 2 units and assessment occasions as level 1 units nested within the level 2 units. For example, consider the data for three individuals in Figure 11.3, where the outcome variable at a given occasion is a behavioral count, namely, the number of alcoholic drinks consumed in the past 2 weeks. This behavior is measured at each of four occasions: the beginning of the first, second, third, and fourth years of high school. All three individuals show a linear change in alcohol consumption over time, and all three individuals have roughly the same average level of drinking across the four time periods. However, the individuals vary in the sharpness of the increase in alcohol consumption across time, as reflected by the differing slopes of the lines plotting their behavioral trajectories. Individual A shows only a slight increase in drinking, Individual B shows a moderate increase in drinking, and Individual C shows a more marked increase in drinking across time. The individuals with the steeper slopes progress toward higher levels of drinking more quickly than the individual with the flatter slope. Instead of explaining individual differences in the average drinking score across time, the focus of this multilevel model might be instead on explaining individual differences in the slope of the line that describes how drinking changes over time. Such "growth curve" analyses are becoming increasingly common in the social sciences. For example, in education research, there often is interest in how reading or math skills change over time and the kinds of trajectories that children show as they move from kindergarten through the elementary school grades. Thus, another option for building theories is to focus on behavioral trajectories over time.

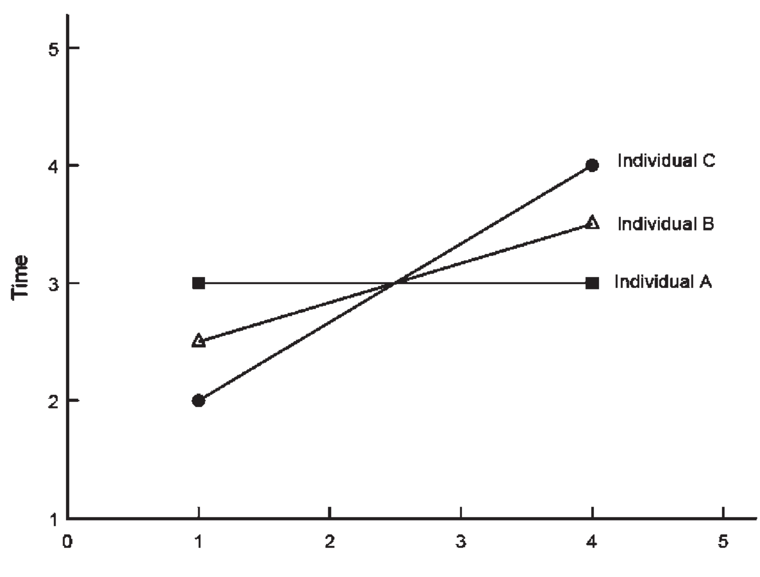


FIGURE 11.3. Example of the analysis of behavioral trajectories.

BOX 11.1. Collaboration

Theory construction often benefits from collaborating with other individuals. Through the exchange of ideas and perspectives, it is possible for individuals involved in the collaboration to gain insights that they may not have obtained working on their own. When we think of collaboration, we usually think of collaborations between scientists, between professors and students, or between students. But collaborative partnerships can be broader than this and can include collaborations between a social scientist, on the one hand, and journalists, policymakers, representatives from institutions or agencies, and key individuals in professional organizations, on the other hand.

Levine and Moreland (2004) have reviewed the research literature on effective collaborations and state that collaborations tend to be more successful if they include both similar and dissimilar orientations of participants. You don't want a collaborator who merely echoes what you think. Instead you want someone who can add to your knowledge base and who can complement the points of view that you bring to the collaboration. On the other hand, if the collaborator has a very different way of characterizing matters, it may be difficult for you to communicate effectively with him or her, as neither of you can transcend the specialized jargon ingrained in your training.

Usually, effective collaborations happen with individuals who are similar in terms of life stage, status, values, and interaction style (Levine & Moreland, 2004). By contrast, similarity on dimensions such as knowledge and abilities inhibits creativity, as it narrows the pooled knowledge base (Farrell, 2001). Farrell (2001) found that more successful collaborations tend to happen in dyads or triads. With a large number of collaborators, usually a natural pairing off of smaller dyads occurs, and it is in these dyads where much of the creative work happens (Farrell, 2001).

When brainstorming ideas as part of a larger group, several disadvantages have been noted. These include evaluation apprehension, production blocking, and a reluctance to discuss unshared ideas. These obstacles can be overcome by talking about them at the outset of the collaboration, and it usually is easier to do so in dyads, where each participant feels comfortable with the other person. The key is to create the right environment for intellectual exchange.

Within a dyad, some people collaborate by writing text together, side-by-side at a word processor, with one person doing the typing. Debates not only focus on ideas but also on such matters as word choice and sentence structure. Alternatively, one person might write an initial draft and then this and subsequent drafts are exchanged, with each person providing feedback to the other. Whichever approach is used, the key is to develop a communal orientation to theory construction rather than one of just exchanging ideas in a linear, noncommunicative way. In general, it is important that both members have a sense of

cont.

ownership of the ideas and a sense that a true collaboration has occurred. This outcome fosters sustained collaboration.

An important part of collaboration is not just generating ideas but also identifying ideas that are good enough for further consideration in the research. Being open to, as well as being willing to give, constructive criticism is important in this regard. Joseph Bordogna, a deputy director of the National Science Foundation, emphasizes the importance of (1) building trust among partners, (2) making sure everyone has something to gain from the collaboration, and (3) ensuring a diversity of perspectives.

Person-Centered Theorizing

Whereas variable-centered approaches to theorizing focus on identifying relationships between variables, person-centered approaches emphasize the identification of groups or "clusters" of individuals who share particular attributes in common. The statistical methodological technique most often associated with person-centered theorizing is cluster analysis and its variants. Cluster analysis applies statistical algorithms to measures of constructs to identify "clusters" of individuals who exhibit common profiles across those measures. As an example, developmental science has identified numerous key parenting dimensions, including (1) expressions of warmth and affection, (2) exertion of control versus laxness, (3) engagement with one's child in shared activities, (4) use of reasoning and explanation during disciplining, (5) use of monitoring and supervision, and (6) quality of communication (Schaefer & DiGeron, 2000). If one conceptualizes parents as being "low," "medium," or "high" on each dimension, there are 3^6 or 729 possible parenting styles defined by the combination of the six dimensions with three levels each. In practice, not all of these styles occur with equal frequency and indeed, some combinations may not occur at all. Cluster analysis is a method that quickly and efficiently identifies parents who share common profiles across the six dimensions and thereby permits scientists to identify the most commonly occurring parenting styles in a population. Such strategies can be applied to any area of study where multiple variables are of interest. For example, there have been many person-centered analyses of health behavior to identify health-based lifestyles of the United States population as well as healthy and unhealthy lifestyles in other countries (Clatworthy, Buick, Hankins, Weinman, & Horne, 2005). Similarly, person-centered analyses have been used to identify clusters of social groupings in social network analysis (Freeman, 2006).

In marketing, cluster analysis is often used to identify market segments for purposes of developing tailored advertising strategies. Segmentation in marketing typically is pursued using four variable categories (Kotler, Roberto, & Lee, 2002): (1) demographic segmentation, which divides populations into segments based on variables such as age, gender, income, occupation, education, religion, ethnicity, and cohort (e.g., generation Y, echo boomers); (2) geographic segmentation, which divides populations into segments according to geographical areas, such as states, regions, counties, cities, and neighborhoods as well as related elements, such as commuting patterns, places of work, and proximity to landmarks; (3) psychographic segmentation, which divides populations into segments based on social class orientation, personality, values, and lifestyle; and (4) behavioral segmentation, which divides populations into segments based on knowledge, attitudes, and practices relevant to the product being marketed (e.g., user status, usage rate, loyalty status, readiness for change). Segments are defined based on some combination of these variable classes. Cluster analysis is used to help marketers identify large population segments that are likely to be receptive to a product, that can be reached in cost-effective ways, and that can be targeted using a common rather than diverse advertising strategy.

As you theorize, consider if a person-centered perspective advances your efforts. Are there multiple dimensions or groups of variables where it would help to identify segments of individuals (or segments of other entities, such as families or organizations) who share common profiles across those dimensions? If so, what profiles or pattern of clusters across the dimensions do you expect to observe and why? Which clusters or profiles do you think will occur most frequently in your population? Why? As you begin to think about "segmenting" your population as a function of response patterns across variables, you may gain new insights into the phenomena you are addressing. For introductions to cluster analysis, see Everitt, Landau, and Leese (2001) and Kaufman and Rousseeuw (2005). For a discussion of recent extensions of cluster analysis to procedures that integrate person-centered and variable-centered theorizing, called mixture modeling, see Lubke and Muthen (2005), Muthen (2001), and Muthen and Muthen (2000).

SUMMARY AND CONCLUDING COMMENTS

We have considered over a dozen different perspectives for thinking about human behavior, including materialism, structuralism, functionalism, symbolic interactionism, evolutionary perspectives, postmodernism, neural networks, systems theory, stage theories, reinforcement theories, humanism, multilevel modeling, and person-centered theorizing. These frameworks complement and augment the frameworks of causal thinking, mathematical modeling, simulations, and grounded theory described in Chapters 7–10, as well as the strategies discussed in Chapter 4 for idea generation. We use the frameworks presented in this chapter by role-playing a scientist who embraces the particular system of thought, and then thinking about how that scientist would analyze the phenomenon in which we are interested. Sometimes this leads to new and productive insights, and sometimes it does not. But by forcing ourselves to invoke each of the perspectives, it helps us think through our phenomena in more depth and often yields insights we would not have thought of otherwise. Of course, there will be some phenomena for which a given framework simply does not seem relevant, and there is no reason to force the matter. But usually we find that a couple of the frameworks are both applicable and useful.

SUGGESTED READINGS

Abbott, A., & Alexander, J. C. (2004). *Methods of discovery: Heuristics for the social sciences*. New York: Norton.—This book has a chapter that adopts an approach similar to the present chapter in which readers are encouraged to think about phenomena from different major schools of thought. However, Abbott focuses on different frameworks and takes a dialectical approach in which one school of thought is seen as in opposition to another school of thought.

Abdi, H., Valentin, D., & Edelman, M. (1999). *Neural networks*. Thousand Oaks, CA: Sage.—An introduction to neural network theories.

Blumer, H. (1969). *Symbolic interactionism: Perspective and method*. Berkeley: University of California Press.—A classic book by a leading expert on symbolic interactionism.

Buss, D., Haselton, M., Shackelford, T., Bleske, A., & Wakefield, J. (1998). Adaptations, exaptations, and spandrels. *American Psychologist, 53*, 533–548.—A general introduction to evolutionary perspectives in the social sciences.

Everitt, B., Landau, S., & Leese, M. (2001). *Cluster analysis*. New York: Hodder-Arnold.—An introduction to the statistical technique of cluster analysis, which forms the cornerstone of many person-centered theory construction approaches.

Gergen, K. (2001). Psychological science in a postmodern context. *American Psychologist, 56*, 803–813.—A central article for the application of postmodern perspectives to the social sciences, with a focus on psychology.

Klein, K., Tossi, H., & Cannella, A. (1999). Multilevel theory building: Barriers, benefits and new developments. *Academy of Management Review, 24*, 243–248.—A special issue on multilevel modeling theories in organizational studies.

Lickleiter, R., & Honeycutt, H. (2003). Developmental dynamics: Toward a biologically plausible evolutionary psychology. *Psychological Bulletin, 129*, 819–835.—An insightful critique of evolutionary perspectives.

McGee, J., & Warms, R. (2003). *Anthropological theory: An introductory history*. Boston: McGraw-Hill.—An introduction to such systems of thought as materialism and structuralism.

Merton, R. (1968). *Social theory and social structure*. New York: Free Press.—A classic on structuralist–functionalist thinking.

Parsons, T. (1971). *The system of modern societies*. Englewood Cliffs, NJ: Prentice-Hall.—Another classic on structuralist thinking.

Raudenbush, S., & Bryk, A. (2001). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.—A insightful and useful discussion of multilevel analysis from the perspective of mathematical and statistical modeling.

Seligman, M., & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychologist, 55*, 5–14.—The lead article of a special issue on positive psychology. The collection of articles also discusses humanistic perspectives more generally.

Skinner, B. F. (1965). *Science and human behavior*. New York: Free Press.—A classic on behaviorism, a major reinforcement theory in psychology.

Weinstein, N., Rothman, A., & Sutton, S. (1998). Stage theories of health behavior: Conceptual and methodological issues. *Health Psychology, 17*, 290–299.—A discussion of the core properties of stage theories.

KEY TERMS

materialism (p. 296)
structuralism (p. 298)
surface structure (p. 299)
deep structure (p. 299)
functionalism (p. 299)
symbolic interactionism (p. 301)
evolution (p. 302)
inheritance (p. 302)
selection (p. 302)
by-products of adaptation (p. 303)
ultimate causation (p. 303)
postmodernism (p. 304)
modernism (p. 305)
deconstruction (p. 306)
neural networks (p. 307)
neuron (p. 307)
input layer (p. 307)
output layer (p. 307)
hidden layer (p. 308)
threshold unit (p. 308)
threshold value (p. 308)
activation index (p. 308)
system (p. 310)
dynamic system (p. 310)
closed system (p. 310)
equilibrium (p. 310)
equifinality (p. 311)
life-cycle model of change (p. 312)
teleological model of change (p. 312)
dialectical model of change (p. 312)
evolutionary model of change (p. 312)
stage theory (p. 313)
developmental synchrony (p. 313)
partial developmental priority (p. 313)
complete developmental priority (p. 313)
reinforcement theories (p. 314)
stimulus (p. 314)
response (p. 315)
eliciting stimulus (p. 315)
discriminative stimulus (p. 315)

reinforcing stimulus (p. 315)
drive (p. 315)
Thurstone's law of effect (p. 315)
positive reinforcers (p. 315)
negative reinforcers (p. 315)
punishment (p. 315)
continuous reinforcement schedule (p. 315)
intermittent reinforcement schedule (p. 315)
stimulus generalization (p. 316)
discrimination (p. 316)
extinction (p. 316)
humanism (p. 317)
focus-oriented strengths (p. 317)
balance-oriented strengths (p. 317)
multilevel model (p. 318)
cluster analysis (p. 323)
mixture modeling (p. 324)

EXERCISES

Exercises to Reinforce Concepts

1. Describe the basic tenets and principles of materialism.
2. Describe the basic tenets and principles of structuralism.
3. Describe the basic tenets and principles of evolutionary perspectives.
4. Describe the basic tenets and principles of postmodernism
5. Describe the basic tenets and principles of neural networks.
6. Describe the basic tenets and principles of systems theories.
7. Describe the basic tenets and principles of stage theories.
8. Describe the basic tenets and principles of reinforcement theories.
9. Describe the basic tenets and principles of humanism.
10. Describe the basic tenets and principles of multilevel modeling.
11. Describe the basic tenets and principles of person-centered theorizing.

Exercises to Apply Concepts

1. Pick an existing theory in an area of interest to you and try to recast it using at least one of the frameworks described in this chapter.
2. Pick an outcome variable and analyze it from the perspective of any two of the frameworks described in this chapter. Compare and contrast the approaches and the conclusions you make.

Part IV
CONCLUDING ISSUES

12

Reading and Writing about Theories

The ability to express an idea is well nigh as important as the idea itself.

—BERNARD BARUCH (1942)

Throughout your career as a scientist, you will read research reports that describe theories. Many of you will write articles that summarize or describe your own theories or the theories of others. In this chapter we discuss practical issues to consider when you read and write about theories. We focus first on reading theories and then on writing about theories.

The ways in which theories are written in professional reports differs by discipline. In disciplines that emphasize experimentation and empirical efforts to test theories, the theories often appear in reports of the results of empirical tests. In disciplines that emphasize emergent/grounded theories, the theories are usually written in narrative form and typically are not the subject of a study designed to formally test an a priori theory. Rather, the theory emerges from the data. There are, of course, exceptions to both of these characterizations. In our discussion of reading about theories, we separate the two approaches, describing first how you will typically see theories presented in outlets emphasizing formal theory tests, and then considering how theories are written about in outlets emphasizing grounded and emergent theorizing. We encourage readers to work through both of these sections no matter what your personal orientations toward theory construction. Our discussion of principles to consider when writing your own reports that present theories focuses on issues that generalize across approaches to theory construction.

READING ABOUT THEORIES IN OUTLETS EMPHASIZING THEORY TESTS AND CONFIRMATORY APPROACHES TO SCIENCE

Journal articles are probably the most common source of material about theories and tests of theories. These articles typically contain four major sections: introduction, methods, results, and discussion. In disciplines that emphasize grounded and emergent theories, the format differs from this, but even in these cases, the spirit of the four sections appears in most articles. We consider each section and how to read and extract from all of them information about a theory. We do *not* consider methodological matters, such as research design and how to evaluate the quality of empirical tests of theories. We focus instead on how to identify and clarify the theory being tested as well as the revisions of the theory the researcher proposes based on the results of the study. It may seem unusual to consider methods for extracting theories from research reports, because one would think that articles would be clear about the theories being considered. Unfortunately, this is not always the case.

We adopt a variable-centered and causal modeling perspective in this section, because these are the dominant approaches used and because it is easier to make certain points. As we have argued in other chapters, even grounded/emergent theorists who primarily rely on process-oriented accounts of explanation sometimes use variable-centered frameworks when analyzing individual phases of their process, so there is something to be gained for such individuals in these sections as well.

The Introduction Section

The introduction section describes the general problem, reviews the relevant literature on the problem, develops the theory to be tested, reviews the relevant literature on the theory, and presents the hypotheses to be tested by the study. Statements also are made about how the research will advance knowledge about the problem area as well as how the research will advance knowledge about the theory.

The essence of theories in most such reports is their variables and the posited relationships between them. A useful strategy for mapping out the theory being tested is to first make a list of all the major variables the author mentions in the introduction section that involve the theory being tested. After doing so, write out the formal conceptual definition of each variable/concept. Sometimes the author will provide an explicit conceptual definition, but other times, the conceptual definition is assumed to be known because the variable or concept is used so often in the scientific literature that there is widespread consensus about its definition. In such cases, you might still write out the conceptual definition so that you can be explicit about the theory, but that is a matter of choice. If the author does not provide a conceptual definition, and you are not aware of a consensual definition, then generate your own "working definition" based on your reading of the article thus far and your past knowledge of the problem area.

Once the concepts/variables and definitions are in place, draw a path diagram of the causal relationships between the variables based on the information in the introduction. Use the methods for path diagramming described in Chapter 7. The path diagram might include direct causal relationships, indirect causal relationships (with either partial or complete mediation), moderated causal relationships, reciprocal causation, spurious relationships, and/or unanalyzed relationships. As you draw the diagram, you may be surprised to find that the theorist does not specify causal links that you think should be addressed. Or you may find that the theorist is vague about certain relationships. As we discuss later, you may still be able to "complete the theory" based on material in other sections of the report. The idea is to complete the path diagram as best you can, based on the material in the introduction.

The Method Section

In the method section the researcher describes methodological features of the empirical study that was conducted to test the theory. This typically includes subsections that describe the research participants, the measures used in the study, how the data were collected, and any other procedural facet that is scientifically relevant. The subsection on the characteristics of the study participants is important theoretically, because it suggests the population to which the theory is applicable. To be sure, the author may envision the theory as applying to populations broader than the one reflected by the particular sample studied, but at the very least, the sampled population provides some sense of the generalizability of the theorizing.

The section on measures also is of interest. It is here that the researcher provides concrete instantiations of the constructs being studied. If the researcher was vague about a conceptual definition in the introduction section, here you can examine it and formulate a conceptual definition based on the measures, because the measures usually are specific and concrete. For example, a researcher might theorize about the construct of intelligence in the introduction section, but never define it. In the method section you discover that the researcher measured intelligence using the Peabody Picture Vocabulary Test (PPVT). As it turns out, the PPVT emphasizes the verbal aspects of intelligence and focuses on the breadth of vocabulary and facility with words. The use of this measure implies a certain conceptual commitment to the meaning of intelligence, and, in this case, the conceptual definition might be construed as one that reflects verbal intelligence.

Sometimes you will be surprised at the way in which a construct is discussed in the introduction section as compared to the instantiation of it that appears in the method section. The measure may reflect a narrower conceptualization than you think is appropriate, or it may reflect a broader conceptualization than what you expected. For example, when discussing the concept of intelligence in the introduction, the researcher might use it in a context that reflects more than verbal intelligence, but when examining the measure, you might discover that the PPVT was used.

Next to each variable you listed in the introduction section, modify any conceptual definitions you had initially written based on your reading of the method section. Then write a brief description of the measure that was used for each construct (or the strategy that was used to manipulate it) next to the conceptual definition. Revisit the concept as it was presented in the introduction, the conceptual definition written next to it, and the measure that was used to assess it. Based on these, you should be able to derive a reasonably clear sense of the variables involved, their conceptual meanings, and how reasonably the measures reflect or represent those meanings.

The Results Section

The results section typically reports how the collected data were analyzed and the ensuing results. Although the analysis may be purely qualitative, in most cases, results sections describe the application of statistical techniques. If a researcher was vague or fuzzy about relationships between variables in the introduction section, it is here that he or she must be more explicit. Almost all major statistical methods focus on characterizing relationships between variables. Just as measures are more specific instantiations of variables, statistical tests are more specific instantiations of presumed relationships between variables. Appendix 12A describes how different statistical tests map onto different causal models and how statistical tests can be used to infer the causal models being addressed.

The Discussion Section

The discussion section addresses, among other things, whether the empirical tests were consistent with the theory. If the theory was not supported, then revisions in the theory are suggested. If the theory was supported, then the researcher often highlights the implications of the results and what future research is needed. Typically, the researcher will encourage the future study of new direct causes, mediators, moderators, extensions to new outcomes, or applying the theory to other contexts and populations to establish generalizability. In all cases, you should be able to repeat the process you performed in the introduction section: Make a list of the variables involved, write out the conceptual definitions of each, and then draw the path diagram to make explicit the relationships between the variables.

In sum, an effective strategy for reading about theories is to make a list of the variables involved in the theory, write out their conceptual definitions, and then draw a path diagram to reflect the presumed causal relationships that operate between the variables. If the theorist is vague or unclear about the above, then clarity can usually be achieved by examining more closely the measures used in the study (the methods section) and the statistical analyses that were pursued (the results section).

READING ABOUT THEORIES IN OUTLETS EMPHASIZING GROUNDED/EMERGENT THEORY

Articles that publish reports of grounded or emergent theories have a somewhat different format than articles based on confirmatory approaches that report theory tests. Although some grounded/emergent theory articles represent a blending of qualitative and quantitative approaches, our discussion here elaborates traditional grounded/emergent theory styles of presentation, with an emphasis on qualitative data.

Articles using grounded/emergent theory typically begin with a statement of the problem, brief background material to provide a context, and a brief characterization of the relevant past literature. The data, typically in the form of field notes and responses to interviews, are woven into the presentation of the theory and the theoretical propositions, as they unfold sequentially in the main body of the article that follows this introductory material. Prior to embarking on this portion, it is common to provide an overview of the major conclusions, so the reader can keep the "big picture" in mind as the particulars are developed. Also, prior to the major theory/data section, there is usually a short method section that discusses in global terms how the data were collected, who they were collected on, and how they were analyzed. It is here that authors build a case that they involved themselves adequately for purposes of conducting an informed grounded/emergent analysis. The article typically ends with a section on conclusions and theoretical/practical implications. It is in this section that the new theory is positioned, relative to extant theory and prior literatures, though this also takes place in the primary narrative as well.

The main body of the article that describes the theoretical propositions and data usually presents a concept or theoretical proposition in the abstract, then provides one or two examples from the collected data to illustrate. The examples often are vivid and image provoking. Here is an excerpt from a paper (Goldin & Jaccard, 2006) by an anthropologist discussing the need to take into account the general living conditions of women when developing programs to address parent–adolescent communication about HIV issues for women living in some of the poorest neighborhoods of Guatemala City, Guatemala:

> People in precarious communities are deeply aware of the links between the high levels of domestic violence and the breakdown of the family—manifested by a lack of communication between parents and children, neglect of children due to tremendous economic pressures, machismo, and a high incidence of single mothers facing difficult situations—and the tendency of young people to look for solace in drugs, alcohol, sexual relationships and gangs (Goldin, Rosenbaum, & Eggleston, 2006). In addition to domestic violence, issues of child care are prominent in women's descriptions of the problems they face. Accounts of children being sold, given up to the care of others, working at young ages, or taking care of yet younger siblings while the mother works, are common. The lack of support networks coupled with the need to work long hours for minimal wages force many women to leave their children unsupervised for long hours every day. The women often show remarkable determination to "manage" otherwise impossible situations. A not untypical account told to us by one woman, Amelia, is illustrative:
>
> > *"It was 1983 and I was very sad. Everything seemed the same to me. I just didn't care. That year my boyfriend had been killed. Leaving a party, a man shot him. I was with him and saw the man who shot him, but there was nothing I could do about it. That was very difficult for me, and so I decided to move to La Esperanza with my mother and siblings. We made a hut with pieces of wood and sheet metal. I met the father of my children here. He was a good friend. He even came to the cemetery with me to bring flowers to my dead boyfriend, and he didn't drink at the time. I became pregnant and soon he started drinking. We didn't get married and he did not live with me. I had my first child and was alone with the baby. What hurt me the most was that he wasn't interested in the baby. He used to live just a block away, but neither*

> *he, nor anyone in his family ever came to visit my baby. It made me so sad. Then a few months later I got pregnant again. I also spent this pregnancy by myself. My daughter was already three months old when we finally started living together. But often I think it would have been better to live alone. I left him once. That time, he wanted to hit me and I said 'no, you are not going to beat me because, just imagine, I feed you, I wash your clothes, and support you, and I'm not going to tolerate your blows.' I left for a month. I went to Escuintla to live with my sister for a while. But then I had to return, because [the father of my children] and his mother had rented out the small place where we lived. He has never worked. I think in all these years, he's worked continuously for about two months. He not only doesn't bring anything home but he takes things that I buy. I have no water, I have no electricity, but he doesn't help. One time, he left for two weeks because he had taken my new iron to sell it and buy liquor. I was so angry that I insulted him. I usually don't swear, but that day I told him all sorts of ugly things and he left. And I had just gotten the results of a pregnancy test. They were positive. I was expecting my third child. I was desperate. He returned later, but it's always the same. The children didn't use to be afraid of him but now, when he's drunk, they don't want to get close to him. He is very rough with the kids and speaks to them in a crude language."*

> While poverty affects all segments of society, women and children tend to be the most affected as they are usually the most vulnerable sector of society, subject to several layers of oppression and marginalization, particularly in societies where machismo prevails. Our research in precarious neighborhoods has made it clear that to ignore the conditions under which these women and children live and to assume that simple messages of the dangers of unsafe sex will resonate with them given all that they must contend with is unrealistic. Consider the words of another mother:

> *"I am not an exemplary mother. I just don't have the time. I don't have time to take care of them. I have to leave them unsupervised most of the time, but I would like to be able to spend more time with them. At least I provide them with food. If I limit the hours I work I would have limited food to give them. Maybe here is where I am failing. I put more effort in the job than in them. They are my life. They need me more. But I feel I give them everything I can."*

> What is needed is an integrated approach that is aimed at the level of the household. As women acquire the skills they need to provide for themselves, the coping strategies they need to help deal with the domestic and neighborhood violence that surrounds them, and as they build better support networks and child care scenarios, they can be more effective at addressing the developmental needs of their adolescent children beyond the basics of survival. Health messages delivered in the context of an organization that has contributed positively to the lives of women in these ways and that has come to earn the trust and respect of the women and the community can be effective.

The paper discussed more generally the development of a women's co-op to address the economic and health obstacles women face in their lives. The paper identified the issues that need to be addressed when structuring women-centered community organizations in precarious neighborhoods to encourage parent–adolescent communication to reduce the HIV infection rate, using narratives and field notes to highlight these points. One conclusion of the paper was that these women have so much with which to contend in their lives, just trying to survive from one day to the next, that until these most basic of needs are addressed, health programs that often conflict with the day-to-day survival strategies of the women and that focus on longer-term health consequences were destined to have marginal impact.

As with any scientific report, one expects the emergent theory to be clearly developed and articulated. In cases where the emergent theory is variable-centered, one can use the same principles described earlier to clarify the constructs and their interrelationships. For process-oriented theories, one can consider creating a process map, as discussed in Chapter 10. As with the variable-centered approach, it often is useful to write out the key concepts and propositions in the theory. The method section of articles using grounded or emergent theory often characterize the elicitation questions for the life and labor histories as well as the focused interviews. The description of these questions may provide clues about conceptual definitions if the researcher was vague about them in the introduction. For accounts that adopt an approach of elaborating propositions and supporting arguments, one can make a list of each conclusion and the arguments relevant to it and then apply the Toulmin model described in Chapter 10 to evaluate the theoretical statements. One also can be sensitive to the 17 argumentation fallacies discussed in Chapter 10.

WRITING ABOUT THEORIES

In this section we identify issues to consider when presenting your theory, focusing on general points that are relevant in all reports. Later sections discuss issues specific to writing in certain outlets.

How You Say It Can Be as Important as What You Say

Over the course of our careers, we have seen articles by colleagues with very good ideas be rejected for publication, and we have seen articles with what we thought were weak ideas published in journals that are highly competitive. Although there are many reasons for this variability, one reason is how the theory is "packaged" in the written product—that is, how the theory is presented. A description of a theory is not unlike the telling of a story, with some people being better storytellers than others. We wish that the world was structured such that it was purely the quality of the idea that mattered. But it is not. If you can't communicate your ideas well, and if you can't get people excited about your ideas, then you are going to have a difficult time publishing your work. You need to be both clear and engaging as you present your ideas.

In graduate school, one of the authors (Jaccard) was taken aside by a senior graduate student who, somewhat tongue in cheek, decided to tell the struggling first-year student the secret to writing scientifically. "Try to think of the most boring and dry way you can say something in the fewest words possible, and you will be a successful scientific writer." In essence, the message was to get to the point and to be concise in getting there. The graduate students were taught to avoid "cute titles" for articles and instead to include the main variables in the title and not much more than that. We also were taught to avoid journalistic tricks, such as starting an article with a gripping, real-life event of an individual who had experienced the phenomenon (or something related to it), and then using this as a lead into the presentation of the science. The objective was to get to the science right away. The strategy of giving phenomena a memorable label (e.g., "fundamental attribution error") also was viewed as "marketing" and was frowned upon.

Increasingly, articles are published today that attempt to engage readers on nonscientific grounds, using clever witticisms to convey basic ideas. Sternberg (2003) and others (e.g., Peter & Olson, 1983) suggest a mindset for scientific writing that is similar to an advertiser: "Keep in mind that you have something to sell, namely your ideas, and sell it" (p. 22). Scientists are only human, and if they have to listen to someone tell a story, they would rather hear it from a good storyteller than a bad storyteller. One hopes that the true science does not get lost in the style of storytelling. If you use writing styles and strategies that are not central to the science so as to engage the reader, do not forget that the science is your first priority!

Briefer Is Better, But Don't Be Too Brief

Readers appreciate papers that are concise and to the point. A lengthy theoretical description is often greeted with dread and sometimes hostility. On the other hand, you need to make your case and provide background to your theory. Don't be afraid to use the space you need, just make sure you need it. Due to the costs of production, most journals have strict limits on the number of manuscript pages that can be published. You typically will find your hands tied because of this restriction. Sometimes you may elect to publish in an outlet not only because the outlet reaches the intended audience, but because it also does not have strict limits on the number of pages. The journal may be less prestigious, but at least you will be able to say what needs to be said and build your case effectively. As the social sciences move more toward electronic media and paperless journals, perhaps editorial boards will be more open to longer articles that develop the underlying logic and implications of a theory in more scholarly ways. Some editors would still want page limits in such cases, arguing that authors need to get to the point. The bottom line is that you need to be scholarly and thorough while at the same time being as brief and concise as possible.

Prepare an Outline

Many people benefit by preparing an outline of the section of the manuscript where the theory is presented (and for that matter, the entire article) prior to actually writing about it. An outline helps you keep the logical sequence of your presentation in mind as you write. It also makes it easier to recognize if you have omitted something crucial. Writing from an outline can help prevent the inclusion of irrelevant thoughts. Some people like to write brief outlines consisting of only key terms or phrases; others prefer to write complete-sentence outlines.

As discussed in Chapter 10, some scientists frame their theories in terms of conclusions or theses and the discussion of arguments for or against those conclusions. One type of outline could follow the basic structure of Toulmin's model of argumentation, as described in Chapter 10. The outline structure would cover (1) an introduction to the problem and a statement of the thesis; (2) presentation of the data/evidence in support of the thesis, along with relevant warrants; (3) development of backings for the warrants; (4) consideration of counterarguments to the thesis; (5) consideration of qualifiers that limit the scope or applicability of the thesis; and (6) a summary restatement of the thesis and its implications.

Provide a Road Map

It often is useful to provide readers at the outset with a "road map" of where you are headed in the narrative. This usually consists of a short paragraph, strategically placed after some introductory orientation, like "In this article, we first discuss the prevalence of adolescent drug use. Next, we consider. . . ." In other words, provide an overview of the structure of the theoretical presentation. It also helps to make liberal use of headings and to make the headings reasonably descriptive.

Provide a Succinct Review of the Current Knowledge

It goes without saying that you need to review past research and summarize the current knowledge about the topic area you are addressing. Often, the "kiss of death" for a paper submitted for publication is the omission of a key article or result from the literature. Journal space is costly and a reader's attention span has its limits, so you usually do not have the luxury of writing about all relevant past research in depth. You might do so in a dissertation, but not in a journal article. If a large body of literature already exists on your topic, try to cite and incorporate published literature reviews. The primary objective of your literature review is to provide a sense of what is already known about the topic you are addressing so as to set the stage for describing how your theory will make a contribution relative to this body of work.

Discuss the Implications and Importance of Your Theory

The importance and implications of your theory may be clear to you, but this does not mean that your readers will automatically recognize them. It helps to be explicit about what new insights and perspectives your theory has to offer. Directly answer the question "What is new here?" and envision a reader who is constantly saying "So, who cares, anyway?" Consider adding a section at a strategic location (e.g., the end of the introduction), tilted "Summary, Innovations, and Implications."

Keep Your Target Audience in Mind

Before writing a paper, it helps to have made a decision, at least tentatively, about the journal to which you plan to submit the paper for possible publication. Various criteria can be used as the basis for selecting a journal. One criterion is to publish in the journal(s) generally acknowledged to be the most rigorous. Another criterion is the audience you want to reach (i.e., the readers of a particular journal). When writing, keep in mind the kinds of scientists (and, when relevant, other key individuals) who will be reading the article, their backgrounds and orientations, and the biases they are likely to hold. This awareness helps you structure and frame your theory in an optimal manner. The type of scientist who reads a journal can best be determined by examining who publishes in the journal, the type of articles published in it, and who cites work published in that journal. The former can be determined by examining recent issues of the journal, and the latter can be determined by referring to the Social Science Citation Index (SSCI, available in libraries or online). The SSCI provides a list of every published article that has cited a target article. In addition, senior researchers and colleagues in the field can be consulted about these matters.

Before a target audience will ever see a paper, however, it must first be accepted for publication. This means that you must also write with another audience in mind, namely, the likely reviewers of the article. If your theory is well articulated, clearly laid out, and makes a contribution, then these strengths will count a great deal toward your paper being accepted by a reviewer. With a complex theory and a complex study (or set of studies) surrounding a theory, it sometimes is difficult to anticipate all the reactions and issues that two or three diverse (and anonymous) reviewers will have. Having a draft of a paper reviewed by your colleagues for purposes of feedback can help in this regard.

Using Figures

Journal space is limited, so editors typically discourage the use of many diagrams or figures. If any, articles generally contain only two or three figures. For variable-centered frameworks that rely on causality, a path diagram (see Chapter 7) can speak a thousand words and is an effective visual aid when presenting a theory. Some theorists provide the path diagram early in the introduction section and then use it to organize an ensuing narrative that considers each path (or a cluster of paths) in the diagram. The relevant literature is reviewed for each path to provide a sense of current knowledge about it and then the contributions of the study to be reported are developed relative to this literature. Other theorists present a narrative organized in this way, but reserve the presentation of the formal diagram until the end of the narrative, as a kind of grand, multivariate summary.

Some theorists list theoretical propositions and label them with phrases like "Proposition 1." Such propositions formalize a theory and highlight its most important points. One can translate a path diagram into propositions and present the logic verbally rather than using a figure. For example, in the case of mediated relationships, a form of syllogistic reasoning is implied with a minor premise, a major premise, and a conclusion. Consider the case in which the impact that watching violence on television has on aggression is assumed to be mediated by the perceived legitimacy of acting aggressively. The syllogistic structure underlying this concept is:

> **Proposition 1:** The more televised violence that people view, the more legitimate they perceive it is to act aggressively.
>
> **Proposition 2:** The more legitimate viewers perceive aggression to be, the more aggressively they behave.
>
> **Proposition 3:** The more televised violence that people watch, the more they will behave aggressively.

In the language of syllogisms, proposition 1 is the minor premise, proposition 2 is the major premise, and proposition 3 is the conclusion. Some scientists prefer presenting theoretical propositions in the above format, whereas others prefer narratives combined with figures. There is no correct way to present a theory in this respect.

Cite Sources for Your Ideas, Text, and Related Items

Section 3.1.3 of the Council of Science Editors' White Paper on Promoting Integrity in Scientific Journal Publications defines plagiarism as "the use of text or other items (figures, images, tables) without permission or acknowledgment of the source of these materials" (p. 39). All of us are familiar with plagiarism. Many are not as familiar with "piracy," which the White Paper defines as "the appropriation of ideas, data, or methods from others without adequate permission or acknowledgment. The intent is the untruthful portrayal of the ideas or methods as one's own" (p. 39). In other words, not only is it unethical to use the exact words of another author without permission, it also is unethical to use ideas originated by others without adequate permission or acknowledgment.

Both authors, as well as a number of our colleagues, have been subject to blatant idea theft and, in some instances, plagiarism as well, and it is frustrating. Although many scientific societies and professional organizations have codes of ethics prohibiting plagiarism, a lesser number have corrective mechanisms for handling the problem. All too often, there is not much one can do about plagiarism or piracy.

That said, sometimes it is hard to remember the sources of your ideas. Moreover, there are instances (as with introductory texts, including this one) where providing citation after citation would burden readers. Further, some journals place limits on the number of citations one can use, so that authors submitting work to such journals sometimes are left in a quandary as to which prior works to cite and which to ignore. So, piracy is not always a cut-and-dry matter. However, you should always approach your writings in the spirit of giving credit where credit is due.

Spelling, Grammar, Typos, and Punctuation

If your manuscript has spelling errors, poor grammar, and/or "typos," then some readers will conclude that you are "sloppy" and don't care enough about your topic. Scientists are noted for being careful and methodical thinkers, and the view is that

BOX 12.1. PowerPoint Presentations of Theories

PowerPoint presentations are commonplace, and you often will present a theory using this form of media. Here is a list of 40 things to consider as you prepare a PowerPoint presentation:

1. Make your first or second slides an outline of your presentation.
2. Follow the order of your outline for the rest of the presentation.
3. Use one or two slides per minute of your presentation.
4. Write in point form, not complete sentences.
5. Include no more than 4–5 points per slide.
6. Avoid wordiness: Use key words and phrases only.
7. If possible, show 1 point at a time by adding points dynamically to the same slide:
   - This helps the audience concentrate on what you are saying.
   - This prevents the audience from reading ahead.
8. Do not use distracting animation.
9. Use at least an 18-point font.
10. Use different-sized fonts for main points than for secondary points.
11. Use a standard font such as Times New Roman or Arial.
12. Place words in all capitals only when necessary—it is difficult to read.
13. Use a color of font that contrasts sharply with the background.
14. Use color to reinforce the logic of your structure (e.g., light blue title and dark blue text).
15. Use color to emphasize a point, but only occasionally.
16. Using color for decoration is distracting.
17. Use backgrounds that are attractive but simple.
18. Use backgrounds that are light.
19. Use the same background consistently throughout your presentation.
20. Data in graphs are easier to comprehend and retain than are raw data.
21. Always title your graphs.
22. Minor gridlines on graphs usually are unnecessary.
23. Proof your slides for spelling mistakes, the use of repeated words, and grammatical errors.
24. If your presentation is not in your first language, have a native speaker check it.
25. Use a strong closing and summarize the main points of your presentation.
26. Consider ending your presentation with a “question slide” that invites your audience to ask questions or that provides a visual aid during the question period.
27. Show up early for your talk. Check whether your equipment works properly.
28. Check whether the projector’s resolution is the same as your laptop’s. If

cont.

it isn't, then your slides may be cropped, may jump, or may lose scan lines.

29. Don't leave Standby Power Management on your laptop on; make sure that your laptop does not turn off if you're inactive for a while during your talk.
30. Don't leave your screen saver on.
31. Don't use the mouse as a pointer. Moving a mouse on a slide show may cause a pointer to appear that is suboptimal in terms of performance.
32. Don't use the edges of the slide. Some projectors crop slides.
33. Do not assume your presentation will work on another person's laptop. Disk failures, software version mismatches, lack of disk space, low memory, and many other factors can prevent this. Check these out before your presentation.
34. Practice moving forward and backward within your presentation. Audiences often ask to see the previous screen again.
35. If possible, preview your slides on the screen you'll be using for your presentation. Make sure that they are readable from the back-row seats.
36. Have a Plan B in the event of technical difficulties (e.g., transparencies and handouts).
37. Practice with someone who has never seen your presentation. Ask him or her for honest feedback about colors, content, and any effects or graphics you've included.
38. Do not read from your slides.
39. Do not speak to your slides. Face the audience, not the slides.
40. When possible, run your presentation from a hard disk rather than a floppy disk or a flash drive. Using a floppy disk or flash drive may slow your presentation.

Note. Compiled from the following websites: *www.iasted.org/conferences/formatting/Presentations-Tips.ppt*; *www.anandnatrajan.com/FAQs/powerpoint.html*; and *kinesiology.boisestate.edu/kines442/tips_for_making_effective_powerp.htm*

these attributes should generalize to other areas of the scientific process, even to the level of spelling, grammar, typos, and punctuation. It is best to be compulsive in this regard.

In sum, when presenting your theory, good communication is the key. How you say something can be just as important as what you say. Being brief and to the point is preferred, but not at the expense of being scholarly. Many theorists benefit by creating outlines prior to writing. If you are developing many ideas, be sure to provide an overview of what you will be covering and make liberal use of headings. State the general problem and then do a succinct review of current knowledge. In addition to presenting the theory, be sure to discuss its implications and importance. As you do so, keep your target audience and reviewers in mind, give credit where credit is due, and correct those typos! In the final analysis, the best way to get a sense of writing styles is to read firsthand articles in the outlets where you will be publishing your work. It is through such readings that you will get a sense of the organizational structures and writing styles that typify successful writing in the areas of study you pursue.

GRANT PROPOSALS, TECHNICAL REPORTS, AND PRESENTATIONS

Social scientists write for different outlets, although by far the most common one is scientific journals. All the principles discussed above will usually serve you well independent of the outlet for which you are writing. Technical reports usually include an "executive summary" that is intended to capture the essence and main conclusions of the larger project in one to three pages. The idea is that a top-level executive usually is too busy to read about the details: He or she just wants to get to the bottom line quickly and efficiently—but have the entire report available should he or she desire to read in greater detail.

It is becoming more common for researchers to seek funding for their research efforts. Grants can be pursued either from federal or state governments or from private, not-for-profit organizations. Typically, the social scientist writes a formal grant proposal and submits it for review by the agency that ultimately decides to fund (or not fund) the research. The level of detail and the description of the underlying theories guiding the research vary considerably, depending on the funding source and the goals. Many agencies focus on applied problems and are most interested in addressing those rather than advancing science or helping to accumulate knowledge about the problem area. In short, their focus is on solutions. Other agencies understand the importance of building a strong knowledge base through both theory and research and demand that strong theories guide efforts of the research they fund. If you pursue funding for your research, look carefully at the proposal guidelines developed by the funding agency, determine the focus and goals of the agency, and try to find examples of successful proposals in your field to see how theory was presented in those proposals.

In terms of oral presentations, you typically will give presentations that are either 15 minutes long (e.g., at a scientific convention) or 45–50 minutes long (e.g., at a job talk or a colloquium). Usually only a small portion of this time is used to describe your theory, perhaps one-fourth or one-third of the allocated time. In oral presentations you might spend a few minutes on a literature review that summarizes current knowledge, a few minutes laying out the theory itself, a few minutes describing what is new and innovative about the theory you propose, and a few minutes on its implications. The book by

Alley (2003) in the Suggested Readings section provides numerous useful strategies for structuring presentations.

SUMMARY AND CONCLUDING COMMENTS

When reading theories in scientific reports, we want to capture the essence of the theory being addressed. For variable-centered theories, a useful strategy is to make a list of the variables in the theory, write out their conceptual definitions, and then draw a path diagram to reflect the presumed causal relationships that operate between the variables. A well-specified theory will clearly articulate the concepts on which it focuses, the nature of those concepts, and the relationships between variables. If the theorist is vague or unclear about these matters, you often will find clarity as the researcher instantiates his or her theory in the methods and results sections. For process-oriented theories, you should list the relevant processes and try to characterize each, perhaps using the process map described in Chapter 10.

When writing about your theory, there are several key points to keep in mind. These include:

1. Attend not only to what you say but also to how you say it.
2. Be brief and to the point, but not at the expense of good scholarship.
3. Work from outlines.
4. Provide readers with an overview of the organization of the paper.
5. Make liberal use of headings.
6. Provide a succinct review of the literature and characterize the current state of knowledge about the phenomena you are studying.
7. Discuss the implications and importance of your theory.
8. Give credit for ideas where credit is due.
9. Always keep in mind the target audience and reviewers.

The best way to get a sense of good scientific writing is to read articles in journals where you intend to publish and take note of the styles used in articles that resonate with you.

SUGGESTED READINGS

Alley, M. (2003). *The craft of scientific presentations.* New York: Springer.—A host of strategies for making effective scientific presentations, based on the techniques of scientists who are effective presenters.

Becker, H., & Richards, P. (2007). *Writing for social scientists: How to start and finish your thesis, book, or article*. Chicago: University of Chicago Press.

Council of Science Editors. (2008). *White paper on promoting integrity in scientific journal publications* (*www.councilscienceeditors.org/editorialpolicies/whitepaper*).—An excellent source of information on the roles and responsibilities of authors, editors, reviewers, sponsoring societies, and media in regard to publishing scientific papers.

Friedland, A., & Felt, C. (2000). *Writing successful science proposals.* New Haven, CT: Yale University Press.—Strategies for writing grant proposals.

Locke, L., Silverman, S., & Spirduso, W. (2004). *Reading and understanding research.* Thousand Oaks, CA: Sage.

Matthews, J., Bowen, J., & Matthews, R. (2000). *Successful scientific writing.* New York: Cambridge University Press.—More writing strategies, but written for the biomedical and medical sciences.

Peter, J. P., & Olson, J. (1983). Is science marketing? *Journal of Marketing, 47,* 111–125.—A discussion of the importance of selling readers on your ideas.

Sternberg, R. (2003). *The psychologists' companion: A guide to scientific writing for students and researchers.* New York: Cambridge University Press.—A book filled with ideas for more effective writing by social scientists.

KEY TERMS

Introduction section (p. 332)
Results section (p. 334)
Methods section (p. 333)
Discussion section (p. 335)

EXERCISES

Exercises to Reinforce Concepts

1. Describe the strategy you would use to discern a theory from the introduction section of a journal article.
2. In what ways can you use the method section to help give clarity to a theory?
3. Describe what you think are the most important points to keep in mind when writing about a theory.

Exercises to Apply Concepts

1. Choose an article that empirically tests a theory and write a short summary of that theory. Identify points in the theory that need clarification or elaboration.
2. Write a report that presents either a theory of your own or an existing theory from the literature using all the principles discussed in this chapter.
3. Prepare a PowerPoint presentation of a theory and present it to someone.

Appendix 12A

Inferring Theoretical Relationships from the Choice of Statistical Tests

This appendix describes how to discern presumed theoretical relationships between variables based on the statistical methods that were chosen by the theorist to analyze data. In doing so, we adopt a causal framework as our point of analysis. We assume you are familiar with each statistical test we discuss. If not, then skip the section describing that particular test.

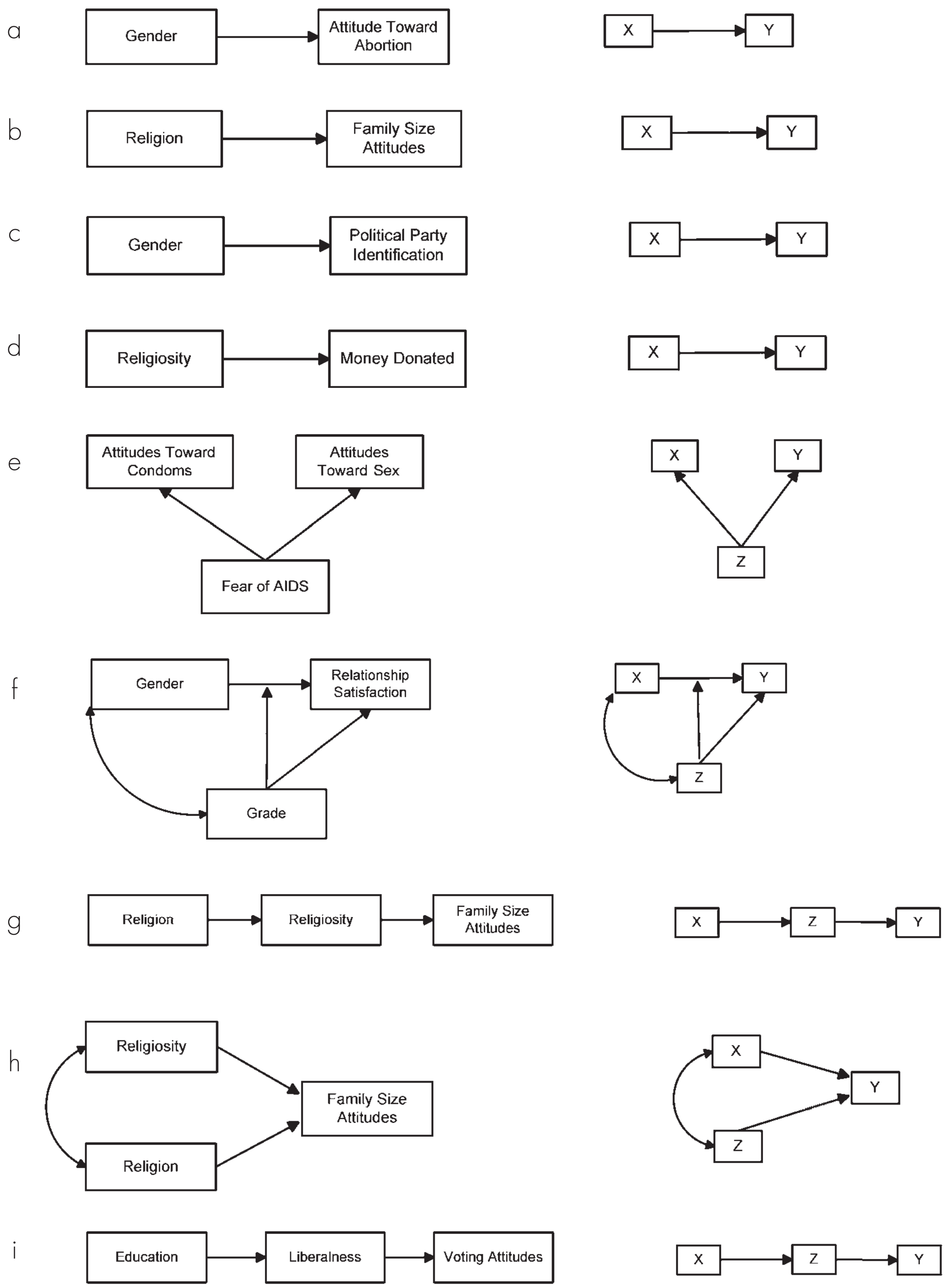
t-TEST

A *t*-test examines a quantitative outcome variable and whether the mean scores on that outcome differ between two groups. For example, one might want to determine if there are gender differences in attitudes toward restricting the legalization of abortion. The two groups, males and females, represent an independent variable or a presumed "cause," and the variable whose means are computed and compared (attitudes toward restricting the legalization of abortions) is the presumed "effect." The underlying causal model that typically motivates the analysis is a direct causal relationship, as shown in Figure 12.1a. This is true for both the independent group's *t*-test and the correlated group's *t*-test. The direction of the mean difference isolates the nature of the relationship, and the magnitude of the effect (in raw score units) is the magnitude of the absolute mean difference. Cohen's *d* is typically used to index the strength of the relationship in a standardized metric. Thus, if a researcher applies an independent group's *t*-test to data or a correlated group's *t*-test, the model in Figure 12.1a is probably the theory that motivates the analysis.

ONE-WAY ANALYSIS OF VARIANCE

The one-way analysis of variance is an extension of the *t*-test, and it, too, has an underlying causal model that usually reflects a direct causal relationship. The different groups comprising the factor represent the independent variable or the presumed "cause," and the variable whose means are compared is the presumed "effect." The underlying causal model that motivates the analysis is shown in Figure 12.1b.

For example, a one-way analysis of variance might be performed to compare three religious groups, Catholics, Protestants, and Jews, on their attitude toward having large families. The overall *F*-test reveals whether there is a relationship between religion (the independent variable) and family size attitudes (the outcome variable). If the overall or omnibus *F*-test is statistically signifi-



cont.

FIGURE 12.1. Causal models underlying statistical tests (text example on left, generic form on right). (a) Two Group/Condition *t*-Test; (b) One-Way Analysis of Variance; (c) Chi-Square Test of Independence and Test of Proportions; (d) Pearson Correlation/Linear Regression: Direct Cause Model; (e) Pearson Correlation: Common Cause or Spurious Effect Model; (f) Two-Factor Analysis of Variance; (g) One-Way Analysis of Covariance: Mediation; (h) One-Way Analysis of Covariance: Independent Influence and Error Reduction; (i) Partial Correlation: Mediation.

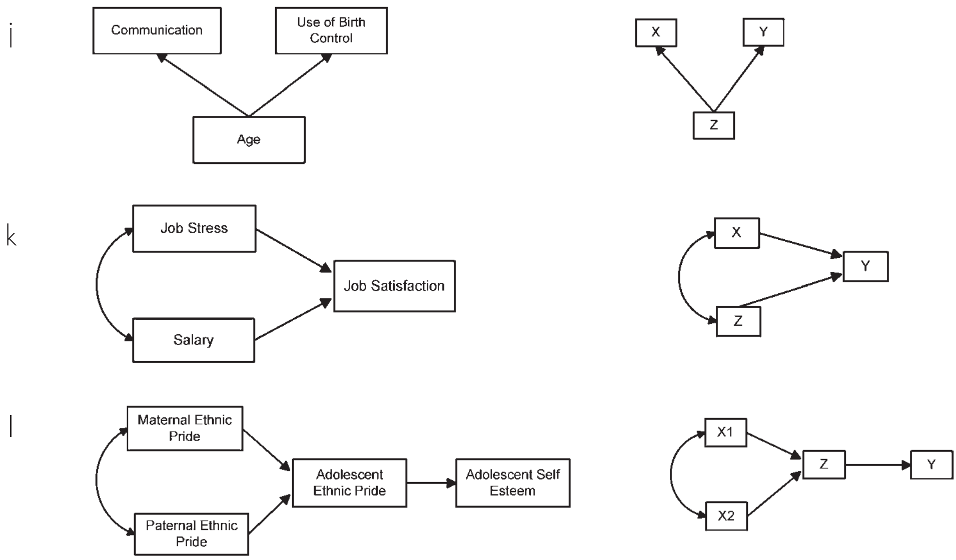


FIGURE 12.1. *(cont.)* (j) Partial Correlation: Common Cause or Spurious Effect Model; (k) Multiple Regression; (l) Hierarchical Multiple Regression—Mediation.

cant, then we conclude there is such a relationship (at least, in terms of means). If the test is not statistically significant, then we cannot conclude that such a relationship exists. Pairwise comparisons of group means isolate the nature of the relationship, and indices such as omega squared or eta squared measure the strength of the relationship in standardized metrics. This underlying causal model holds for both between-groups and repeated measures analysis of variance. Thus, if a researcher applies a one-way analysis of variance to data, the model in Figure 12.1b is probably the theory that motivates the analysis.

CHI-SQUARE TEST OF INDEPENDENCE AND TESTS OF PROPORTIONS

The chi-square test of independence examines the association between two categorical variables. For example, a theorist might believe that there are gender differences in political party identification and test this hypothesis using the chi-square test of independence. Gender is the independent variable or the presumed "cause," and political party identification is the presumed "effect." A statistically significant omnibus chi-square test is consistent with the proposition that a relationship exists. Figure 12.1c presents the causal model that motivates this analysis. Follow-up comparisons focused on specific cells of the contingency table isolate the nature of the relationship, and indices such as Cramer's *V* or the phi coefficient measure the strength of the relationship on a standardized metric.

A variant of the chi-square test of independence is a test of the difference between two proportions. For example, a researcher might compare the proportion of male seventh graders who have engaged in sexual intercourse with the proportion of female seventh graders who have engaged in sexual intercourse. The significance test roughly maps onto the chi-square test of independence in a 2 × 2 contingency table, with gender as rows and whether the individual has had sex as columns. Thus, the model in Figure 12.1c also captures the causal model that typically motivates a test of the difference between two proportions.

In sum, if a researcher applies a chi-square test of independence or a test of the difference between two proportions to data, the model in Figure 12.1c is probably the theory that motivates the analysis.

PEARSON CORRELATION AND/OR LINEAR REGRESSION

A Pearson correlation examines the degree to which the relationship between two variables approximates a linear relationship. The predictor variable, X, is the presumed "cause," and the criterion variable, Y, is the presumed "effect." For example, a researcher might compute the correlation between individuals' degree of religiosity and how much money they donate to religious organizations over the course of a year. Figure 12.1d presents the causal model that often motivates a correlation analysis, which is a direct causal relationship.

Another model that might motivate the calculation of a correlation coefficient is a spurious effect or "common cause" model. If two variables are thought to be related because they share a common cause, then there should be a correlation between them. For example, if fear of AIDS impacts both attitudes toward using condoms and attitudes toward having sex, then one would expect attitudes toward condoms and attitudes toward having sex to be correlated. A statistically significant correlation between two variables is consistent with the proposition that they share a common cause. The "common cause" or spurious correlation model motivating a correlation coefficient is presented in Figure 12.1e.

There are other models that might motivate a bivariate correlation analysis (e.g., to identify a variable as a potential confound or to document an unanalyzed relationship), but the above two models are by far the most common ones motivating the calculation of a correlation. Usually you can discern which of the two theoretical orientations the researcher has adopted based on material from the introduction and/or discussion section.

FACTORIAL ANALYSIS OF VARIANCE

We describe the case of factorial analysis of variance (FANOVA) using a two-factor design. Extensions to three-factor or four-factor designs are straightforward. In a two-factor analysis of variance, the quantitative dependent variable whose means are being computed and compared is the outcome variable or the presumed "effect." We refer to one of the factors as X and the other factor as Z; these are the presumed "causes." In FANOVA one typically examines three "effects": the main effect of X, the main effect of Z, and the interaction between X and Z. There are separate omnibus F-tests for each effect. As an example, let X be gender, Z be grade in school, and the outcome variable be adolescents' degree of satisfaction with their relationship with their mother. The FANOVA tests if there is a significant mean difference in satisfaction as a function of gender (collapsing across grade in school), a significant mean difference in satisfaction as a function of grade in school (collapsing across gender), and if the gender difference in satisfaction varies significantly as a function of grade in school. Figure 12.1f presents the causal model that typically motivates a FANOVA. The main effects reflect direct causes, and the interaction effect reflects a moderated relationship. A researcher who conducts a factorial analysis of variance usually is testing if the data are consistent with the causal paths noted in Figure 12.1f. This underlying causal model holds for both between-groups and repeated measures of factorial analysis of variance.

ONE-WAY ANALYSIS OF COVARIANCE

A one-way analysis of covariance (ANCOVA) is comparable to a one-way analysis of variance (ANOVA), except that a covariate is included in the analysis. For example, a researcher might analyze the same variables as the one-way ANOVA (described above) by comparing three religious groups (Catholics, Protestants, and Jews) on attitudes toward having large families. However, a covariate is introduced into the analysis, namely, each individual's degree of religiosity. The comparisons are conducted on group differences in attitudes toward large families, holding constant (or covarying out) religiosity.

The underlying causal model for this analytic method usually takes one of two forms. The first is a mediation model and is illustrated in Figure 12.1g. The covariate is thought to be a (complete) mediator of the effect of the "factor" on the dependent variable. For example, the effect of religion on family size attitudes is thought to be mediated by the fact that religious groups differ, on average, on how religious their members are. If, after holding the covariate constant, the effect of the factor on the outcome variable is no longer statistically significant, then this is consistent with the proposition that the covariate is a complete mediator of the effect. If the effect of the factor on the outcome variable remains statistically significant, then this is consistent with the proposition that the factor has independent effects on the outcome variable over and above the covariate.[1] Thus, some form of mediational model often motivates an analysis of covariance.

The second causal model commonly underlying analysis of covariance is shown in Figure 12.1h. In this model, the researcher views the factor and the covariate as each having independent influences on the outcome variable. Including the covariate in the analysis reduces bias in the parameter estimates (by avoiding left-out variable error) and also reduces the error term, which can increase the power of the statistical test. Mediation is not an issue in this model. The focus instead is on assessing the effect of the factor while simultaneously taking into account the effect of the covariate.

Although other causal models can motivate a one-way analysis of covariance (e.g., a spurious effect model), the above two (or a combination of them) are the most common motivators of this test. If a researcher conducts an analysis of covariance, it usually is with one of the above two models in mind. More often than not, you can discern which of the two models the researcher is using from the text in the introduction and discussion sections.

FACTORIAL ANALYSIS OF COVARIANCE

A factorial analysis of covariance is comparable to a factorial analysis of variance, except that a covariate is included in the analysis. The causal model underlying the factorial analysis of covari-

[1]Formal statistical tests to evaluate partial or complete mediation are more complex than simple covariate-based procedures such as the ones described in this section. This does not alter the fact that mediational models often are the motivating sources of covariate-based analyses, such as analysis of covariance, partial correlation, and hierarchical regression.

ance is the same as that for the factorial analysis of variance (see Figure 12.1f), except for the additional role of the covariate. Like the one-way ANCOVA, the covariate is typically seen as taking on the role of a mediator or as an independent influence on the outcome, whose inclusion in the model reduces error and increases the statistical power of the test.

PARTIAL CORRELATION

A partial correlation analysis is similar to a Pearson correlation analysis but includes a covariate. One of two causal models typically motivates a partial correlation analysis. The first model is a mediational model and is illustrated in Figure 12.1i. A researcher might want to test if a variable Z is a partial or complete mediator of the effect of X on Y. For example, to test if liberalness mediates the relationship between education and attitudes toward voting for a Democratic presidential candidate, a partial correlation between years of education and attitudes toward the candidate might be computed, partialing out liberalness. A statistically significant partial correlation is consistent with the proposition that education has an independent effect on attitudes toward voting for the Democratic presidential candidate over and above liberalness. If a statistically significant correlation between education and voting attitudes becomes nonsignificant when liberalness is partialed out, then this is consistent with the proposition that liberalness mediates the relationship between education and voting attitudes (but see footnote 1).

The second causal model that often motivates a partial correlation analysis is a common cause or spurious effect model. For example, a researcher might argue that the relationship between parent–adolescent communication about birth control and adolescent intentions to use birth control is spurious due to the common cause of age of the adolescent: As adolescents get older, parents are more likely to talk with them about issues surrounding the use of birth control. As adolescents get older, they also are more likely to use birth control, often for reasons that have nothing to do with parental communication. So the argument might be that any association between parent–adolescent communication and use of birth control is spurious. If one calculates a partial correlation between parent–adolescent communication and adolescent use of birth control using age of the adolescent as a covariate, a statistically significant correlation is consistent with the proposition that the correlation is not spurious. Figure 12.1j reflects these dynamics.

As before, other types of causal models may motivate the use of partial correlation analysis, but the above two models are the most common bases for partial correlation.

MULTIPLE REGRESSION

Multiple regression analysis involves regressing an outcome variable, Y, onto two or more predictor variables. Figure 12.1k presents the causal model that typically motivates this analytic method. As an example, a researcher might regress job satisfaction onto two variables: (1) how stressful the job is and (2) how much an individual gets paid, as reflected by his or her annual salary. A statistically significant regression coefficient for a given predictor is consistent with the proposition that there is a causal path between the presumed "cause" (the predictor variable) and the presumed "effect" of job satisfaction.

It is possible to use multiple regression to pursue moderated relationships using product terms and a wide variety of other types of casual relationships. However, traditional applications of multiple regression with several predictors evaluate models of the form in Figure 12.1k.

HIERARCHICAL REGRESSION

Hierarchical regression analysis is typically used to address either mediation or covariate control. For the case of mediation, the researcher typically enters the mediators on the first step of the analysis and then the more distal causes on the second step. In the case of complete mediation, the prediction is that the increment in the multiple correlation at the second step will not be statistically significant. A statistically significant increment in the squared multiple correlation implies that one or more of the distal causes has an independent effect on the outcome variable over and above the mediators.

For example, a researcher might study self-esteem as an outcome variable in adolescent Latinos. The researcher theorizes that self-esteem is influenced by ethnic pride (i.e., how much pride the adolescent takes in being Latino). The adolescent's ethnic pride, in turn, is thought to be a function of the ethnic pride of his or her mother and father. The impact of maternal and paternal ethnic pride on adolescent self-esteem is thought to be mediated by adolescent ethnic pride. Figure 12.1l presents the underlying causal model. The analyst performs a hierarchical regression by first regressing adolescent self-esteem onto adolescent ethnic pride. A statistically significant regression coefficient should be observed. If more than one mediator is thought to be operating, then all the mediators are entered into the regression equation at this first step. At the second step, the researcher adds the two distal variables to the equation, hypothesizing that the regression coefficients for the two distal variables will not be statistically significant, but for the mediators, the regression coefficients will remain statistically significant.[2]

For the case of covariate control, the covariates that are to be controlled are entered on the first step of the analysis and then the key "causes" of interest are entered on the second step. If the key causes impact the outcome over and above the covariates, then the increment in the multiple correlation should be statistically significant. If a significant increase in the multiple correlation is observed, then this is consistent with the proposition that there is an effect of the "causes" on the outcome variable over and above the covariates. The fundamental causal model for this analysis is the same as for multiple regression (see Figure 12.1k), with the covariates representing one set of predictors and the key "causes" representing the other set of predictors, all in the same equation.

As with all the methods discussed, one can usually discern the underlying model motivating a hierarchical regression analysis from the introduction and discussion sections. There are other models that can motivate hierarchical regression, but the ones described above are those you will most typically encounter.

LOGISTIC REGRESSION AND THE GENERALIZED LINEAR MODEL

Logistic regression is analogous to multiple regression but focuses on the case where the outcome variable is dichotomous. The generalized linear model extends regression-like analyses to

[2]There are better ways of testing the mediation model than hierarchical regression; these make use of structural equation modeling. Nevertheless, you occasionally will encounter use of hierarchical regression to test mediation.

count variables (e.g., Poisson regression; negative binomial regression), categorical outcomes with more than two categories (e.g., multinomial logistic regression), or ordinal outcomes (e.g., ordinal regression). The fundamental causal model that motivates all these techniques is the same as that depicted in Figure 12.1k for multiple regression. When hierarchical tests are conducted in the context of these models, causal models similar to those discussed for hierarchical regression typically motivate the analysis.

STRUCTURAL EQUATION MODELING

Structural equation modeling (SEM), an analytic method designed to evaluate the viability of causal models, can be drawn in the form of path diagrams. The theoretical structure the researcher is using when SEM methods are pursued is usually evident from the accompanying path diagrams.

There are other statistical tests you will encounter, but the above represent a good sampling of the major statistical tests used by social scientists. Once an investigator applies one or more of these tests, he or she usually is invoking a causal model of the form that we described above for each test. If a researcher has not made clear the presumed theoretical relationships between variables or the underlying causal model that is the focus, then you can often discern this by examining the statistical tests that the researcher reports in the results section.

13

Epilogue

The most important fundamental laws and facts of physical science have all been discovered, and these are now so firmly established that the possibility of their ever being supplemented in consequence by new discoveries is exceedingly remote.

—ALBERT MICHELSON, noted German scientist (1903), stated before Einstein's groundbreaking work in physics

This final chapter covers odds and ends that did not fit well in previous chapters. The organization is eclectic, but we felt it important to touch on the issues. We begin by discussing competing theories. We then consider post hoc theorizing, wherein one derives theory after data have been collected rather than before. Next, we revisit the notion of influential theorizing and describe some of the career implications of pursuing creative theorizing. Finally, we recap areas for self study that build theory construction skills.

COMPETING THEORIES

As you build a theory about a phenomenon, you may develop competing theories that either make opposite predictions or that account for the same phenomena but use different explanations and assumptions. In such situations, you should not abandon the competing theories and force yourself to endorse a single theory. Rather you should embrace the differences and pursue research that competitively tests the theories against one another (or conduct research that will allow you to integrate them). Many scientists consider research that conclusively chooses between two or more logical and plausible theories to be inherently more interesting than a study that yields results regarding a single theory (Platt, 1964).

POST HOC THEORIZING

In some disciplines, scientists scorn what they call *post hoc theorizing*. The confirmatory approach to research is one wherein a scientist uses a theory to derive a set of testable hypotheses and then designs a study to test those hypotheses. Based on the test, the theory is either rejected or not rejected. In post hoc theorizing, a researcher collects data, sometimes based on an initial set of ideas that are only vaguely articulated, and then uses statistical analysis to examine the many possible relationships in the data, noting cases where relationships manifest themselves. The researcher then creates a theory to account for the observed relationships. The theory is post hoc in the sense that it is derived after data are collected and is based on inspection of statistical indices of many possible relationships between variables. The major complaints against this approach are that (1) it capitalizes on chance relationships in the data and (2) that any semi-creative scientist can generate an explanation after the fact for just about anything, making the theory arbitrary.

Despite these objections, some camps in the social sciences are sympathetic to post hoc theorizing or variants of it. As noted in Chapter 10, social scientists who adopt a grounded or emergent theory orientation are open to letting theory emerge from data rather than positing theories a priori. Such theorists do not necessarily embrace "statistical witch-hunts" across large numbers of statistical associations, but the general philosophy accepts post hoc analysis. There also is a branch of statistics entrenched in Bayesian principles that does not evaluate the worth of an explanation based on when the explanation is offered (namely, either a priori or post hoc). According to Bayesians, what matters most is whether the data are consistent or inconsistent with the explanation. More specifically, the crucial factors to consider are (1) the prior odds of the truth of the explanation before the data were collected; (2) the probability that the observed data would pattern themselves in the way they did, given that the explanation is true; and (3) the probability that the observed data would pattern themselves in the way they did, given that the explanation is not true. The timing of when the explanation is offered (a priori or post hoc) is only relevant if it affects one or more of these factors. For a discussion of Bayesian logic, see Lee (2004).

Our own position is that if one adequately takes into account chance effects (and there is a host of strategies for doing so), then theories offered on a post hoc basis are worthy of consideration.

INFLUENTIAL SCIENCE

As noted in Chapter 4, some theories represent major contributions to the field, whereas other theories make modest or small contributions. Most theoretical advances are of this latter type. One way that scientists measure the impact of a published article is to examine its citation count—that is, how many other published articles cite that article. The idea is that the more people who cite an article, the more influential it is. By this standard, most articles published in the social sciences are not very influential. The modal citation count for a published article is zero, and this includes self-citations. About 6% of publishing scientists produce 50% of all publications, a figure that has not changed much since the 1920s. The reality is that it is difficult to publish impactful theories and research, and relatively few scientists do so.

Highly cited theories and research are not necessarily creative. Some of the most widely cited articles in the social sciences are straightforward psychometric studies that develop a measure for a heavily researched construct. We are not downplaying the importance of such measurement-oriented research, but on a dimension of creativity and theoretical insight, the contributions of such articles are more limited.

Nor is creative research necessarily impactful. John Cacioppo (2004, p. 118) comments on recent trends in the field of social psychology as follows:

> We value the prize of a theory that makes non-obvious predictions, that illuminates flaws in social reasoning and interactions, that illustrates not only the inadequacy but the idiocy of common sense. Such work is unquestionably clever, but does the pursuit of the witty at the expense of the comprehensive put personality and social psychologists at risk of becoming the editorial cartoonists of the social sciences?

Cacioppo's concern derives from the growing number of theories and studies in social psychology that focus on cute, fun, clever, or counterintuitive demonstrations rather than focusing on the nuts and bolts of building comprehensive theories of behavior.

In the context of the above, we reiterate one last time the importance of deciding to be creative and seeking impact in your work. As you pursue your theoretical efforts, we urge you to keep the values of creativity, impact, and comprehensiveness at the forefront.

CAREERS AND CREATIVE THEORIZING IN SCIENCE

Creative research that pursues big ideas can be greeted with resistance and skepticism by scientists entrenched in the dominant scientific paradigms. The work may initially be difficult to publish, and it may take years to build an empirical base in support of the new theory. This is not true of research that accepts standard presuppositions and then builds on scientific knowledge in more modest and traditional ways. If asked, most scientists would not admit to a bias against "crowd-defying" ideas. But the reality is that scientists are trained to think in terms of a common paradigm, and this paradigm dominates their evaluation of other ideas. As such, there inevitably is some bias against crowd-defying ideas. This bias has implications for career pursuit and development.

Consider the traditional path of a student pursuing doctoral training in the social sciences with the idea of securing an academic career in a research-oriented college or university. Typically, a PhD is obtained after 4 years of advanced study. At this time, doctoral students seek to transition to the role of assistant professor via a position at a college or university. When making decisions about which applicants to interview, a major criterion used by search committees is that applicants must set themselves apart from other applicants in terms of the quality and quantity of their scientific publications. A typical position at the assistant professor level will have 75 or more applicants, although this varies depending on the discipline. In some universities and subfields, there may be as many as 300 applicants. Initial decisions are made by a search committee, who review the vitae that applicants submit. Usually the top three or four candidates are invited for in-person interviews. These interviews generally occur within the first few months of the fourth or final year of one's graduate training. This means that research you conduct during your fourth year is less likely to be considered when it comes time to apply for jobs, because this research has not been accepted for publication at the time your vita is considered by the search committee. The time frame for a graduate student to produce publications is, in reality, 3 years, not 4 years (unless one extends graduate training beyond the customary 4-year period).

For a research project to be published, the scientist must develop an idea, collect and analyze data to test or explore the idea, and then write a research report and submit it to a journal for publication. This process takes a minimum of 6 months, usually longer. Once the paper is submitted for publication, it must be peer reviewed (which, depending upon the journal, takes 3 months or more), the result of which usually is an editorial decision to "revise and resubmit" before final acceptance (which takes an additional 3 months). Thus, the minimum time lag between when a research idea is formed and when a research report is accepted for publication typically is about 1 year. This time lag means that only research that is started *before* the early part of the third year of graduate study has a chance of being accepted for publication in time to have an impact on your job prospects. The bottom line is that it is the first 2 years of graduate study that are crucial for deciding one's career prospects, and these need to be research focused. Most students spend their first year of study taking required courses and gaining scientific knowledge and background to better prepare themselves for conducting research. In the typical 4-year program, this leaves the precious second year of graduate study as the one that ultimately has a major impact on one's career fate.

One way that graduate students adapt to these demands is to become involved in the research programs of their major professor. By working in an already established program of research, students can produce joint authored publications with their major professors at a competitive and reasonably rapid pace.

It is against this backdrop that one must face the realities of developing and conducting research related to a crowd-defying idea. Given the possible resistance to such an idea and the time it takes to lay the groundwork and empirical base for such a theory, it is unlikely that a graduate student who pursues this type of research will be competitive on the job market when his or her graduate studies are completed because it is unlikely that this work will be completed in time to be published. Our suggestion is to strike a balance between the idealistic side of science and the practical side of science. By all means, pursue your big ideas. Take the risk and see where it leads you. Science needs this type of creative pursuit. But while you are doing so, also develop a second research program that will lead you to the publications you need, the training you need, and the contacts you need to be competitive on the job market. This will require extra work on your part, but hopefully it will be worth it. We are not suggesting that your "second" research program be something in which you are not interested or something to which you are not committed. Indeed, it often can supplement or be a smaller part of your crowd-defying idea. But in making the choice to purse a big idea, we suggest that you give practical considerations their due weight, or else you may find yourself with no viable prospect for a job in academia. Idealists may not like this advice. It may indeed be possible for you to carve out your own, highly innovative program of research without regard to a second research program of the nature we are describing. But this will be rare and it requires an exceptional student, an exceptional major professor (who is willing to support the student), an exceptional graduate program, and an exceptional set of environmental supports.

These same time demands are equally oppressive for new faculty hired as assistant professors. In our opinion, the pressure is even greater because such faculty can no longer rely upon their major professors, at least in the way they did during graduate school. New professors typically work for 7 years and then are evaluated for tenure. If they are granted tenure, they are given a lifetime contract. If they are not granted tenure, they are fired. If denied tenure, it is difficult to obtain another academic position. So tenure decisions are penultimate for new professors.

Because institutions want to provide you with a year to find a new job should you be denied tenure, the formal tenure review process typically starts at the beginning of your sixth year of employment, not your seventh year. At this time, you are expected to have established an independent research program and, depending on your particular discipline and the type and prestige of your host institution, you are expected to have published anywhere from 5 to 25 high-quality publications in your field. Given the time lags associated with publishing described above, any research you conduct near the end of your fifth year probably will be too delayed to count toward your tenure decision. This means that you have 4–4 ½ years to produce. Your first year as an assistant professor usually is consumed by the logistics of moving to a new location, preparing lectures for the courses you will teach, and getting a research program set up (e.g., recruiting graduate and undergraduate students to help, buying equipment, and so on). This reduces the time you have to conduct research to gain tenure to years 2–4. To make matters worse, the faculty who ultimately evaluate you for tenure start to form impressions of how productive you are from the beginning of your employment, so if you produce nothing in your first year or two, this often leads to a negative impression. It is against this backdrop that you must decide to pursue a crowd-defying idea, with its attendant time delays and risks.

Our advice to young scientists in these circumstances is the same as what we gave the graduate student. Pursue your big ideas. But also develop a second research program that respects the practical constraints in which you are operating. If you do not, you may end up being fired with no prospect of pursuing the big ideas you treasure.

The above scenarios vary from one discipline to the next, but the bottom line is the same. You need to be both an idealist and a realist. If you pursue only the idealistic side of science (the side that we love and that emphasizes advancing knowledge in creative and fulfilling ways), your career may come to a screeching, unexpected halt, and you may find yourself unable to do what you love. If you pursue the practical side only, you may find yourself poorly equipped to pursue an independent research program that makes sufficient contributions to science. Indeed, you may not receive tenure because your research is considered pedestrian. Find a balance between idealism and realism, but never let go of your idealism. Pursue it with a vengeance when reasonable opportunities present themselves.

SCIENTIFIC PARADIGMS

Thomas Kuhn (1962) has written a series of books on scientific paradigms and the structure of scientific revolutions. His work is controversial and has numerous critics, but some of his ideas are worth mentioning here. Kuhn characterizes scientific research as consisting of paradigms that represent commonly agreed-upon assumptions, meanings, and methods for characterizing and summarizing phenomena. According to Kuhn, the evolution of a paradigm is influenced by many factors, and true science is just one of them. Historical, social, and political factors also come into play.

Kuhn (1962, 1970, 2000) described scientific revolutions as passing through four stages. During the first stage, called *normal science*, ordinary scientific research and scientific progress occurs. The dominant paradigm sets the agenda for the type of theorizing and research that is pursued, and it shapes scientific thinking accordingly. The second stage is the *appearance of an anomaly*. Some problems are harder to solve than others by the prevailing paradigm, and this difficulty can induce a period of crisis. Or flaws in the dominant paradigm start to accumulate, leading some scientists to question it. The third stage is *crisis*. If a problem or finding provokes a *crisis*, the grip of the paradigm on science weakens. A few especially creative scientists break out of the confines of the dominant paradigm, rejecting one or more of its defining tenets and proposing new ones. The final stage is *revolution*, where the advocates of the emerging paradigm gain control of the power structure in science, namely, journal editorships, granting agencies, and textbooks. The old paradigm is replaced by the new, beginning the cycle again with that of normal science.

According to Kuhn (1970), scientists are trained to adhere to the dominant paradigm to the point that science is reduced to small incremental gains in knowledge that resist extra-paradigmatic thinking. This research increases the precision and scope of thinking, but all within the constraints of the presuppositions of the day. Kuhn (1970) argues that some scientific decisions are not logical but instead are based on values, politics, and consensus in the scientific community. He believes that younger researchers are more likely to recognize these problems and show a willingness to explore different approaches because they are less entrenched in the paradigm of the day.

Kuhn's analysis has critics, but certain elements ring true. Being aware of the dominant paradigms and the way in which one's thinking can be channeled by them is a key to breaking through to make truly creative contributions.

A PROGRAM OF SELF-STUDY

Throughout this book we have made reference to a "theoretical toolbox" and identified areas wherein it would be useful for you to pursue further reading over and above your substantive domains of interest. The areas are diverse, and it is rare to find a graduate program that emphasizes all of them. A summary list of these areas is provided below. It is not that areas omitted from this list are unimportant. Rather, we believe that the list serves as a starting point. In these days of electronic search and access to articles and books, it is much easier to set aside regular time to update oneself in these areas, even if one must do so fleetingly.

We suggest five strategies to stay current on these topics. First, once or twice a year, conduct an electronic search through reference databases in libraries and examine all abstracts of articles and books in a given topic area. Flag a subset to obtain and read. By reading the abstracts and titles, you will gain a sense of "what is out there," even though you may not choose to read every article. Read review articles and/or articles/books that you judge will most impact your substantive work. Our electronic searches include not only major electronic library-based databases (such as Medline, PsychInfo, Sociological Abstracts, Social Work Abstracts), but also major sellers of books (e.g., *amazon.com*). For the latter, when you select a book to examine, many sellers display related books and books that others who purchased that book also purchased. This can facilitate the process of identifying potentially relevant books.

Second, identify a small group of scientists whose work you have come to admire and whom you judge to be particularly creative and insightful. Every year, conduct an electronic search to examine the abstracts of all articles they published that year. Obtain and read a subset of these articles, focusing on those that seem interesting or useful. A variant of this strategy is the identification of web pages of certain individuals or professional organizations (e.g., American Academy of Pediatrics) to regularly consult for information and leads about further study and relevant literatures.

Third, identify certain journals that consistently publish important articles in a given target area. Make a point of scanning the abstracts of every article published in those journals in a given year, again with the idea of then identifying a subset of articles to obtain and read.

Fourth, use several search engines (e.g., Google, Yahoo!) to conduct electronic searches of the areas of study. This strategy taps into web pages that people have created to address different topics and can lead to a wealth of information about a topic area that is not in the formal scientific literature. To be sure, there is a great deal of misinformation on the web, and one must be careful in this regard. But there also is a tremendous amount of useful information to be found.

Fifth, talk to colleagues about their work and attend colloquia sponsored in your department and other departments as much as possible, especially when the topics are on the list provided below. Many researchers also find it useful to attend professional conferences and their workshops.

In no particular order of importance, here we list a suggested program of self-study (that we personally use) of areas of which you might keep abreast over and above your substantive domains of interest:

1. *Creativity*: Read current literatures on creativity that may give you new thinking strategies.
2. *Psychometrics*: Read literatures in psychometrics that help you define and think about constructs in general as well as in your substantive areas of study.
3. *Philosophy of science*: Read the classics in philosophy of science as well as more current-day perspectives on science from the perspective of philosophers.
4. *Causality and causal modeling*: Read literature on statistical, methodological, conceptual, and philosophical issues on causality in general as well as moderation, mediation, spuriousness, reciprocal causality, longitudinal modeling, and latent variables, in particular.
5. *Simulation methods*: Read about simulation methodology, especially as it applies to emerging technologies.
6. *Mathematical modeling*: In addition to mathematical modeling in general, read literatures on chaos theory and catastrophe theory.
7. *Grounded/emergent theory construction*: Read literature on strategies for constructing grounded theories and qualitative data analysis.
8. *Rhetoric and communication theory*: Read literature on methods of argumentation, logic, and theories and methods of debate.
9. *Cognitive biases and limitations of information processing*: This literature tends to appear mostly in cognitive psychology, social psychology, consumer behavior, and decision making. It will help keep you attuned to biases that can enter into your thinking.
10. *Historical methodology*. For research involving analysis of primary sources and archival materials, staying abreast of good historical methodological practices is important.
11. *General systems of thought*: Touch base with the literatures on all of the topics discussed in Chapter 11. These include neural networks, stage theories, reinforcement theories, humanism and positive psychology, symbolic interactionsism, systems theory, materialism, structuralism, functionalism, postmodernism, and evolutionary perspectives.
12. *Statistics*: Keep abreast of recent developments in probability theory, exploratory statistical methods, structural equation modeling, nonlinear modeling, multilevel modeling, and longitudinal methods.
13. *Scientific and nonscientific writing techniques and presentation strategies*: Strive to find more effective ways of presenting your theories.

In addition, to staying abreast of the above literatures, we also make it a point to stay current on our substantive areas of interest and, as time permits, to read biographies and autobiographies, and other works of fiction and nonfiction that may expand our thinking.

CONCLUDING COMMENTS

A great deal of material has been covered that hopefully will help you develop theories or build on existing theory. As you pursue your career and strive to develop theories, we want to underscore the importance of being resilient. You will face criticism and negative feedback. All of us do. The key is not to let this feedback grind your efforts to a halt or get you too discouraged. Listen to it, work with it, decide what is worthwhile, and keep a forward focus. The classic characterization of reaction to feedback by the famous psychologist Benton Underwoood (1957, p. 222) is as much applicable today as it was some 50 years ago:

> The rejection of my own manuscripts has a sordid aftermath: (a) One day of depression; (b) one day of utter contempt for the editor and his accomplices; (c) one day of decrying the conspiracy against letting truth be published; (d) one day of fretful ideas about changing my profession; (e) one day of re-evaluating the manuscript in view of the editor's comments followed by the conclusion that I was lucky it wasn't accepted.

Despite the ups and downs you ultimately will experience, if you stay with it and be resilient, and if you approach your science with curiosity, an open mind, and a diverse theoretical toolbox from which to work, positive contributions probably are in your future.

SUGGESTED READINGS

Kuhn, T. (1996). *The structure of scientific revolutions* (3rd ed.). Chicago: University of Chicago Press.—A classic on paradigm shifts in science, but controversial.

Kuhn, T., & Conant, J. (2000). *The road since structure: Philosophical essays, 1970–1993, with an autobiographical interview*. Chicago: University of Chicago Press.—More on Kuhn's work.

KEY TERMS

post hoc theorizing (p. 355)

stage of normal science (p. 360)

stage of appearance of an anomaly (p. 360)

stage of crisis (p. 360)

stage of revolution (p. 360)

EXERCISES

Exercises to Reinforce Concepts

1. Discuss the reasons why people object to post hoc theorizing. Is post hoc theorizing ever justified?

2. What are the major stages in Kuhn's analysis of paradigm shifts? Characterize each.

Exercise to Apply Concepts

1. Develop a program of self-study for this next semester or year. Implement it!

References

Aarts, H., & Dijksterhuis, A. (2000). Habits as knowledge structures: Automaticity in goal-directed behavior. *Journal of Personality and Social Psychology*, *78*, 53–63.

Abbott, A., & Alexander, J. C. (2004). *Methods of discovery: Heuristics for the social sciences*. New York: Norton.

Abdi, H., Valentin, D., & Edelman, B. (1999). *Neural networks*. Thousand Oaks, CA: Sage.

Achinstein, P. (1983). *The nature of explanation*. New York: Oxford University Press.

Ackoff, R. L. (1991). *Ackoff's fables*. New York: Wiley.

Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior.* Englewood Cliffs, NJ: Prentice-Hall.

Alford, R. (1998). *The craft of inquiry: Theories, methods, evidence*. New York: Oxford University Press.

Alley, M. (2003). *The craft of scientific presentations*. New York: Springer.

Amabile, T. M. (1983). *The social psychology of creativity.* New York: Springer-Verlag.

Anderson, N. H. (1965). Primacy effects in personality impression formation using a generalized order effect paradigm. *Journal of Personality and Social Psychology*, 2, 1–9.

Anderson, N. H. (1981). *Foundations of information integration theory.* New York: Academic Press.

Anderson, N. H. (1991). *Contributions to information integration*. Hillsdale, NJ: Erlbaum.

Anderson, N. H., & Hubert, S. (1963). Effects of concomitant verbal recall on order effects in personality impression formation. *Journal of Verbal Learning and Verbal Behavior*, 2, 379–391.

Arnheim, R. (2004). *Visual thinking*. Berkeley: University of California Press.

Austin, J. T., & Vancouver, J. B. (1996). Goal constructs in psychology: Structure, process, and content. *Psychological Bulletin*, *120*, 338–375.

Babbie, E. R. (1973). *Survey research methods*. Belmont, CA: Wadsworth.

Bachorowski, J., & Newman, J. P. (1990). Impulsive motor behavior: Effects of personality and goal salience. *Journal of Personality and Social Psychology*, *58*, 512–518.

Bacon, S. (2005). Positive psychology's two cultures. *Review of General Psychology*, *9*, 181–192.

Bakeman, R., & Casey, R. L. (1995). Analyzing family interaction: Taking time into account. *Journal of Family Psychology*, *9*, 131–143.

Bancroft, J., & Reinisch, J. (1990). *Adolescence and puberty*. New York: Oxford University Press.

Barber, N. (2008). Evolutionary social science: A new approach to violent crime. *Aggression and Violent Behavior*, *13*, 237–250.

Barron, F., & Harrington, D. (1981). Creativity, intelligence, and personality. *Annual Review of Psychology*, *32*, 439–476.

Barry, A. (1997). *Visual intelligence: Perception, image, and manipulation in visual communication*. Albany: State University of New York Press.

Bassili, J. N., & Brown, R. (2005). Implicit and explicit attitudes: Research, challenges, and theory. In D. Albarracin, B. Johnson, & M. Zanna (Eds.), *The handbook of attitudes* (pp. 543–574). Mahwah, NJ: Erlbaum.

Becker, H. (2003). *Tricks of the trade: How to think about your research while you're doing it*. Chicago: University of Chicago Press.

Bednar, J., & Miikkulainen, R. (2003). Learning innate face preferences. *Neural Computation*, *15*, 1525–1557.

Bem, D. J. (1970). *Beliefs, attitudes, and human affairs*. Belmont, CA: Brooks/Cole.

Bernard, T., Paoline, E., & Pare, P. (2005). General systems theory and criminal justice. *Journal of Criminal Justice*, *33*, 203–211.

Berntson, G. G., & Cacioppo, J. T. (2003). A contemporary perspective on multilevel analyses and social neuroscience. In F. Kessel & P. Rosenfield (Eds.), *Expanding the boundaries of health and social science: Case studies in interdisciplinary innovation* (pp. 18–40). London: Oxford University Press.

Bereczeki, T. (2000). Evolutionary psychology: A new perspective in the behavioral sciences. *European Psychologist*, *5*, 175–190.

Blalock, H. M. (1964). *Causal inferences in non-experimental research*. Chapel Hill: University of North Carolina Press.

Blanton, H., & Jaccard, J. (2006a). Arbitrary metrics in psychology. *American Psychologist*, *61*, 27–41.

Blanton, H., & Jaccard, J. (2006b). Tests of multiplicative models in psychology: A case study using the unified theory of implicit attitudes, stereotypes, self esteem, and self concept. *Psychological Review*, *122*, 155–169.

Blanton, H., & Jaccard, J. (2008). Unconscious prejudice: A concept in pursuit of a measure. *Annual Review of Sociology*, *34*, 277–297.

Blumer, H. (1969). *Symbolic interactionism: Perspective and method*. Berkeley: University of California Press.

Boas, F. (1897). *The social organization and secret societies of the Kwakintl indians*. (Report of the U.S. National Museum for 1895), pp. 311–738.

Bodenhausen, G., & Lambert, A. (2003). *Foundations of social cognition: A Festschrift in honor of Robert S. Wyer, Jr.* Hillsdale, NJ: Erlbaum.

Bohrnstedt, G. W. (1970). Reliability and validity assessment in attitude measurement. In G. F. Summers (Ed.), *Attitude measurement* (pp. 80–99). Chicago: Rand McNally.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, *83*, 7280–7287.

Borsari, B., & Carey, K. B. (2003). Descriptive and injunctive norms in college drinking: A meta-analytic integration. *Journal of Studies in Alcohol*, *64*, 331–341.

Brinberg, D., & Hirschman, E. (1986). Multiple orientations for the conduct of marketing research: An analysis of the academic/practitioner distinction. *Journal of Marketing*, *50*, 161–173.

Brinberg, D., & McGrath, J. (1985). *Validity and the research process*. Newbury Park, CA: Sage.

Brodbeck, M. (1968). *Readings in the philosophy of the social sciences*. New York: Macmillan.

Bronfenbrenner, U. (1986). Ecology of the family as a context for human development: Research perspectives. *Developmental Psychology*, *22*, 723–742.

Brown, T. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.

Bunge, M. (1961). Causality, chance, and law. *American Scientist*, *69*, 432–488.

Buss, D., Haselton, M., Shackelford, T., Bleske, A., & Wakefield, J. (1998). Adaptations, exaptations, and spandrels. *American Psychologist*, *53*, 533–548.

Cacioppo, J. T. (2004). Common sense, intuition, and theory in personality and social psychology. *Personality and Social Psychology Review*, *8*, 114–122.

Cacioppo, J. T., Berntson, G. G., Lorig, T. S., Norris, C. J., Rickett, E., & Nusbaum, H. (2003). Just because you're imaging the brain doesn't mean you can stop using your head: A primer and set of first principles. *Journal of Personality and Social Psychology*, *85*, 650–661.

Carlin, J. (1962). *Lawyers on their own*. New Brunswick, NJ: Rutgers University Press.

Carnap, R. (1936). Testability and meaning. *Philosophy of Science*, *3*, 419–471.

Carnap, R. (1937). Testability and meaning. *Philosophy of Science*, *4*, 1–40.

Carnap, R. (1971). A basic system of inductive logic, Part 1. In R. Carnap & R. C. Jeffrey (Eds.), *Studies in inductive logic and probability* (Vol. I, pp. 18–39). Berkeley: University of California Press.

Carnap, R., & Jeffrey, R. (Eds.). (1971). *Studies in inductive logic and probability* (Vol. I). Berkeley: University of California Press.

Cartwright, N. (2007). *Hunting causes and using them*. New York: Cambridge University Press.

Catalano, R., Berglund, L., Ryan, J., Lonczak, H. S., & Hawkins, J. D. (2002). Positive youth development in the United States: Research findings on evaluations of positive youth development programs. *Prevention and Treatment*, *5*, Article 15. Retrieved April 14, 2003, from *journals.apa.org/prevention/volume5/toc-jun24-02.htm*.

Cederman, L. (2005). Computational models of social forms: Advancing generative process theory. *American Journal of Sociology*, *110*, 864–893.

Chaplin, J. P., & Krawiec, T. S. (1960). *Systems and theories of psychology.* New York: Holt, Rinehart & Winston.

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. Newbury Park, CA: Sage.

Chestnut, R. W., & Jacoby, J. (1980). Product comprehension: The case of permanent versus term life insurance. *Advances in Consumer Research*, *7*, 424–428.

Chia, R., & Tsoukas, H. (2002). On organizational becoming: Rethinking organizational change. *Organization Science*, *13*, 567–582.

Cialdini, R. (2001). *Influence: Science and practice*. Boston: Allyn & Bacon.

Cialdini, R. (2003). Crafting normative messages to protect the environment. *Current Directions in Psychological Science*, *12*, 105–109.

Clatworthy, J., Buick, D., Hankins, M., Weinman, J., & Horne, R. (2005). The use and reporting of cluster analysis in health psychology: A review. *British Journal of Health Psychology*, *10*, 329–358.

Clement, J. (1988). Observed methods for generating analogies in scientific problem solving. *Cognitive Science*, *12*, 563–586.

Clement, J. (2006). Thought experiments and imagery in expert protocols. In L. Magnani (Ed.), *Model-based reasoning in science and engineering* (pp. 1–16). London: King's College.

Clement, J. (2008). *Creative model construction in scientists and students: The role of imagery, analogy, and mental simulation*. Mahwah, NJ: Erlbaum.

Cohen, B. (1956). *Franklin and Newton: An inquiry into speculative Newtonian experimental science and Franklin's work in electricity as an example thereof.* Philadelphia, PA: American Philosophical Society.

Cole, J. W., & Wolf, E. (1974). *The hidden frontier: Ecology and ethnicity in an Alpine valley*. Berkeley: University of California Press.

Coleman, J. (1994). *Foundations of social theory*. Boston: Harvard University Press.

Cook, S. D. N., & Brown, J. S. (1999). Bridging epistemologies: The generative dance between organizational knowledge and organizational knowing. *Organization Science*, *10*, 381–400.

Coombs, C., Dawes, R., & Tversky, A. (1970). *Mathematical psychology: An elementary introduction*. Englewood Cliffs, NJ: Prentice-Hall.

Cooper, R., & Law, J. (1995). Organization: Distal and proximal views. *Research in the Sociology of Organizations*, *13*, 237–274.

Cornelissen, J. (2004). What are we playing at?: Theatre, organization, and the use of metaphor. *Organization Studies*, *25*, 705–726.

Cosentino, M. (2007). *Case in point: Complete case interview preparation*. Needham, MA: Burgee Press.

Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI): Professional manual*. Odessa, FL: Psychological Assessment Resources.

Crano, W., & Burgoon, M. (2002). *Mass media and drug prevention: Classic and contemporary theories and research*. Mahwah, NJ: Erlbaum.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281–302.

Cunningham, J., & Macan, T. (2007). Effects of applicant pregnancy on hiring decisions and interview ratings. *Sex Roles*, *57*, 497–508.

Cylwik, H. (2001). Notes from the field: Emotions of place in the production and interpretation of text. *International Journal of Social Research Methodology*, *4*, 243–250.

Davis, J., Bingham, C., & Eisenhardt, K. (2007). Developing theory through simulation methods. *Academy of Management Review*, *32*, 480–499.

Davis, J., Eisenhardt, K., & Bingham, C. (2005). *Complexity theory, market dynamism, and the strategy of simple rules*. Unpublished manuscript, Stanford Technology Ventures Program, Palo Alto, CA.

Dawkins, R. (1996). *Climbing Mount Improbable*. New York: Norton.

Dawson, R. E. (1963). Simulation in the social sciences. In H. Guetzkow (Ed.), *Simulation in social sciences* (pp. 221–252). Englewood Cliffs, NJ: Prentice-Hall.

De Cock, C., & Sharp, R. (2007). Process theory and research: Exploring the dialectic tension. *Scandinavian Journal of Management*, *23*, 233–250.

Denzin, N. K., & Lincoln, Y. (2000). *Handbook of qualitative research*. Thousand Oaks, CA: Sage.

Deutsch, M., & Krauss, R. (1960). The effect of threat upon interpersonal bargaining. *Journal of Abnormal and Social Psychology*, *61*, 181–189.

Dick, R. (1990). *Convergent interviewing*. Brisbane, Australia: Interchange.

Dittus, P., & Jaccard, J. (2000). The relationship of adolescent perceptions of maternal disapproval of sex and of the mother–adolescent relationship to sexual outcomes. *Journal of Adolescent Health*, *26*, 268–278.

Dixon, J., & Moore, C. (2000). The logic of interpreting evidence of developmental ordering: Strong inference and categorical measures. *Developmental Psychology*, *36*, 826–834.

Dollard, J., & Miller, N. E. (1950). *Personality and psychotherapy*. New York: McGraw-Hill.

Donovan, J., & Jessor, R. (1985). Structure of problem behavior in adolescence and young adulthood. *Journal of Consulting and Clinical Psychology*, *53*, 890–904.

Dubin, R. (1976). Theory building in applied areas. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 17–39). Chicago: Rand McNally.

Duck, S. W. (1982). A topography of relationship disengagement and dissolution. In S. Duck (Ed.), *Personal relationships 4: Dissolving personal relationships* (pp. 112–139). London: Academic Press.

Duncan, T. E., Duncan, S. C., Okut, H., Strycker, L. A. H., & Small, H. (2003). A multilevel contextual model of neighborhood collective efficacy. *American Journal of Community Psychology*, *32*, 245–252.

Durand, M. (2002). *Severe behavior problems: A functional communication training approach*. New York: Guilford Press.

Eagly, A. H., Johannesen-Schmidt, M., & van Engen, M. (2003). Transformational, transactional, and laissez-faire leadership styles: A meta-analysis comparing women and men. *Psychological Bulletin*, *129*, 569–591.

Edin, K., & Kefalas, M. (2005). *Promises I can keep: Why poor women put motherhood before marriage*. Berkeley: University of California Press.

Edwards, R. (1979). *Contested terrain*. New York: Basic Books.

Einstein, A. (1934). *Essays in science*. New York: Philosophical Library.

Eisenhardt, K. E. (1989). Building theories from case study research. *Academy of Management Review*, *14*, 532–550.

Ellet, W. (2007). *The case study handbook: How to read, discuss, and write persuasively about cases*. Boston: Harvard Business School Press.

El-Shamy, S. (2005). *Role play made easy: 25 structured rehearsals for managing problem situations and dealing with difficult people*. New York: Pfeiffer.

Enzensberger, H. M. (1974). A critique of political ecology. *New Left Review*, *84*, 3–32.

Epstein, C. F. (1983). *Women in law*. Garden City, NY: Doubleday.

Everitt, B., Landau, S., & Leese, M. (2001). *Cluster analysis*. New York: Hodder-Arnold.

Farmer, J., & Geanakoplos, J. (2002). *Beyond equilibrium and efficiency*. New York: Oxford University Press.

Farrell, M. P. (2001). Collaborative circles: Friendship dynamics and creative work laboratories to collaboratories: A new organizational form for scientific collaboration. *European Journal of Social Psychology*, *31*, 707–729.

Festinger, L. (1962). *A theory of cognitive dissonance*. Palo Alto, CA: Stanford University Press.

Fishbein, M. (1973). The prediction of behavior from attitudinal variables. In C. D. Mortensen & K. Soreno (Eds.), *Advances in communication research* (pp. 3–31). New York: Harper & Row.

Fishbein, M., & Ajzen, I. (1975). *Beliefs, attitudes, intentions and behavior*. Reading, MA: Addison-Wesley.

Fishbein, M., Jamieson, K., Zimmer, E., von Haeften, I., & Robin, N. (2002). Avoiding the boomerang: Testing the relative effectiveness of anti-drug public service announcements before a national campaign. *American Journal of Public Health*, *92*, 238–245.

Fogel, R. (1964). *Railroads and American economic growth*. Baltimore: Johns Hopkins University Press.

Folger, R., & Turillo, C. J. (1999). Theorizing as the thickness of thin abstraction. *Academy of Management Review*, *24*, 742–758.

Frank, P. (1961). *Modern science and its philosophy*. New York: Collier Books.

Freeman, B. (1993). *Substance abuse treatment: A family systems perspective*. Thousand Oaks, CA: Sage.

Freeman, L. (2006). *The development of social network analysis*. Vancouver: Empirical Press.

Friedman, M. (1996). *Type A behavior: Its diagnosis and treatment*. New York: Plenum Press.

Friedman, M., & Rosenman, R. H. (1974). *Type A behavior and your heart*. New York: Plenum Press.

Fromkin, H. L., & Streufert, S. (1976). Laboratory experimentation. In M. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 415–466). Chicago: Rand McNally.

Fussell, P. (1983). *Class: A guide through the American status system*. New York: Touchstone.

Geake, J., & Gross, M. (2008). Teachers' negative affect toward academically gifted students: An evolutionary psychological study. *Gifted Child Quarterly*, *52*, 217–231.

Gergen, K. J. (1985). The social constructionist movement in modern psychology. *American Psychologist*, *40*, 266–275.

Gergen, K. J. (2001). Psychological science in a postmodern context. *American Psychologist*, *56*, 803–813.

Gergen, M., & Gergen, K. (2003). *Social construction: A reader*. London: Sage.

Gibbons, F., Lane, D. J., Gerrard, M., Pomery, E., & Lautrup, C. (2002). Drinking and driving: A prospective assessment of the relation between risk cognitions and risk behavior. *Risk Decision and Policy*, *7*, 267–283.

Gilbert, N., & Abbott, A. (2005). Introduction. *American Journal of Sociology*, *110*, 859–863.

Glaser, B. (1978). *Theoretical sensitivity: Advances in the methodology of grounded theory*. New York: Sociology Press.

Glaser, B. (1992). *Basics of grounded theory analysis*. Mill Valley, CA: Sociology Press.

Glaser, B., & Strauss, A. (1965). *Awareness of dying*. Chicago: Aldine.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.

Glisson, C. (2002). The organizational context of children's mental health services. *Clinical Child and Family Psychology Review*, *5*, 233–253.
Goffman, E. (1959). *The presentation of self in everyday life*. New York: Anchor Books.
Goffman, E. (1967). *Interaction ritual: Essays on face-to-face behavior*. New York: Doubleday.
Goldin, L. R. (1987). The "peace of the market" in the midst of violence: A symbolic analysis of markets and exchange in Western Guatemala. *Ethnos*, *3*, 368–383.
Goldin, L. R. (2009). *Global Maya: Work and ideology in rural Guatemala*. Tucson: University of Arizona Press.
Goldin, L. R., & Jaccard, J. (2006, July 1). *Poverty and parent outreach for AIDS prevention in the DR*. Unpublished National Institutes of Health grant proposal.
Goldin, L., Rosenbaum, B., & Eggleston, S. (2006). Women's participation in non-goverment organizations: Implications for poverty reduction in precarious settlements in Guatemala City. *City and Society, 18*, 260–287.
Goleman, D. (1995). *Emotional intelligence*. New York: Bantam Books.
Gooding, D. (1992). The procedural turn: Or, why do thought experiments work? In R. Giere (Ed.), *Cognitive models of science* (pp. 45–76). Minneapolis: University of Minnesota Press.
Gorsuch, R. (1983). *Factor analysis*. Hillsdale, NJ: Erlbaum.
Green, P., & Tull, D. S. (1975). *Research for marketing decisions*. Englewood Cliffs, NJ: Prentice-Hall.
Gregory, R. (1997). *Eye and brain*. Princeton, NJ: Princeton University Press.
Guetzkow, H., Akger, C. F., Brody, R. A., Noel, C., & Snyder, R. C. (1963). *Simulation in international relations*. Englewood Cliffs, NJ: Prentice-Hall.
Guilamo-Ramos, V., Jaccard, J., & Dittus, P. (2004). Parental and school correlates of binge drinking in middle school. *American Journal of Public Health*, *95*, 894–899.
Guilamo-Ramos, V., Litardo, H., & Jaccard, J. (2003). Prevention programs for reducing adolescent problem behaviors: Implications of the co-occurrence of problem behaviors in adolescence. *Journal of Adolescent Health*, *36*, 82–86.
Hannerz, U. (1993). *Cultural complexity: Studies in the social organization of meaning*. New York: Columbia University Press.
Hastings, B., & Shaffer, B. (2008). Authoritarianism: The role of threat, evolutionary psychology, and the will to power. *Theory and Psychology*, *18*, 423–440.
Hatfield, A., & Hipel, K. (2002). Risk and systems theory. *Risk Analysis*, *22*, 1043–1057.
Headland, T., Pike, K., & Harris, M. (1990). *Emics and etics: The insider/outsider debate*. Thousand Oaks, CA: Sage.
Higgins, E. T., Rholes, W., & Jones, C. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology*, *13*, 141–154.
Hill, D. (2006). Theory in applied social psychology: Past mistakes and future hopes. *Theory and Psychology*, *16*, 613–640.
Hoffman, D. (2000). *Visual intelligence: How we create what we see*. New York: Norton.
Hoffmann, J. P. (2003). A contextual analysis of differential association, social control, and strain theories of delinquency. *Social Forces*, *81*, 753–785.
Hollander, E. P. (1967). *Principles and methods of social psychology*. Oxford, UK: Oxford University Press.
Hoyer, W. D., & Jacoby, J. (1983). Three-dimensional information acquisition: An application to contraceptive decision making. In R. P. Bagozzi & A. Tybout (Eds.), *Advances in consumer research*, *10*, 618–623.
Hull, C. (1943). *Principles of behavior*. New York: Appleton-Century.
Hume, D. (1975). *An enquiry concerning human understanding* (3rd ed.; P. N. Nidditch, Ed.). Oxford, UK: Clarendon Press. (Original work published 1777)
Hume, L., & Mulcock, J. (2004). *Anthropologists in the field: Cases in participant observation*. New York: Columbia University Press.
Hutchins, E. (1996). *Cognition in the wild*. Cambridge, MA: MIT Press.

Isen, A. (1987). Positive affect, cognitive processes, and social behavior. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 20, pp. 203–253). San Diego: Academic Press.

Jaccard, J. (1974). Predicting social behavior from personality traits. *Journal of Research in Personality*, *7*, 358–367.

Jaccard, J. (1998). *Interaction effects in factorial analysis of variance*. Newbury Park, CA: Sage.

Jaccard, J. (2009). *Unlocking the contraceptive conundrum: Reducing unintended pregnancies in emergent adulthood*. Washington, DC: National Campaign to Prevent Teen and Unplanned Pregnancy.

Jaccard, J., & Blanton, H. (2005). The origins and structure of behavior: Conceptualizing behavior in attitude research. In D. Albarracin, B. Johnson, & M. Zanna (Eds.), *The handbook of attitudes* (pp. 125–171). Mahwah, NJ: Erlbaum.

Jaccard, J., Dodge, T., & Guilamo-Ramos, V. (2005). Meta-cognition, risk behavior, and risk outcomes: The role of perceived intelligence and perceived knowledge. *Health Psychology*, *24*, 161–170.

Jaccard, J., & Turrisi, R. (2003). *Interaction effects in multiple regression*. Newbury Park, CA: Sage.

Jacoby, J. (1975). Perspectives on a consumer information processing research program. *Communication Research*, *2*, 203–215.

Jacoby, J. (1977). Information load and decision quality: Some contested issues. *Journal of Marketing Research*, *14*, 569–573.

Jacoby, J. (1984). Perspectives on information overload. *Journal of Consumer Research*, *10*, 432–435.

Jacoby, J. (2002). Stimulus–organism–response reconsidered: An evolutionary step in modeling (consumer) behavior. *Journal of Consumer Psychology*, *12*, 51–57.

Jacoby, J., & Chestnut, R. (1977). *Amount, type and order of information acquisition in purchasing decisions*. Final report for the National Science Foundation, GI-43687.

Jacoby, J., & Chestnut, R. (1978). *Brand loyalty: Measurement and management*. New York: Wiley.

Jacoby, J., Jaccard, J., Currim, I., Kuss, A., Ansari, A., & Troutman, T. (1994). Tracing the impact of item-by-item information accessing on uncertainty reduction. *Journal of Consumer Research*, *21*, 291–303.

Jacoby, J., Jaccard, J., Kuss, A., Troutman, T., & Mazursky, D. (1987). New directions in behavioral process research: Implications for social psychology. *Journal of Experimental Social Psychology*, *23*, 146–174.

Jacoby, J., Morrin, M., Jaccard, J., Gurhan, Z., & Maheswaran, D. (2002). Attitude formation as a function of incremental information input: A procedure for studying on-line processing models of attitude formation. *Journal of Consumer Psychology*, *12*, 21–34.

Jacoby, J., Speller, D. E., & Berning, C. (1974). Brand choice behavior as a function of information load: Replication and extension. *Journal of Consumer Research*, *1*, 33–42.

Jacoby, J., Speller, D. E., & Kohn, C. (1974). Brand choice behavior as a function of information load. *Journal of Marketing Research*, *11*, 63–69.

James, W. (1890). *The principles of psychology*. New York: Dover.

Jessor, R. (1994). Problem–behavior theory and the life course in adolescence: Epistemology in action. *Medicine and Mind*, *8*, 57–68.

Jessor, R., & Jessor, S. L. (1977). *Problem behavior and psychosocial development: A longitudinal study of youth*. New York: Academic Press.

Johar, G. V., Jeddidi, K., & Jacoby, J. (1997). A varying parameter averaging model of on-line brand evaluations. *Journal of Consumer Research*, *24*, 232–247.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, *3*, 430–454.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*, 237–251.

Kalick, S. M., & Hamilton, T. (1986). The matching hypothesis reexamined. *Journal of Personality and Social Psychology*, *51*, 673–682.

Kaplan, A. (1964). *The conduct of inquiry*. San Francisco: Chandler.

Katz, D. (1960). The functional approach to the study of attitudes, *Public Opinion Quarterly*, *24*, 163–204.

Kaufman, L., & Rousseeuw, P. (2005). *Finding groups in data: An introduction to cluster analysis*. New York: Wiley.

Kazdin, A. (2008). *The Kazdin method for parenting the defiant child*. New York: Houghton Mifflin.

Kelinman, S., & Copp, M. (1993). *Emotions and fieldwork*. Newbury Park, CA: Sage.

Kenny, D. A., Kashy, D., & Cook, W. (2006). *Dyadic data analysis*. New York: Guilford Press.

Kimble, G. A. (1984). Psychology's two cultures. *American Psychologist*, *39*, 833–839.

Kotler, P., Roberto, N., & Lee, N. (2002). *Social marketing: Improving the quality of life*. Thousand Oaks, CA: Sage.

Kowalski, B. (2006). *Computational logic as a dual process model of thought*. London: Imperial College.

Kraska, P. (2004). *Theorizing criminal justice*. Long Grove, IL: Waveland.

Kruglanski, A. W. (2004). The quest for the gist: On challenges of going abstract in social and personality psychology. *Personality and Social Psychology Review*, *8*, 156–163.

Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

Kuhn, T. S. (1970). Logic of discovery or psychology of research? In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 1–22). Cambridge, UK: Cambridge University Press.

Kuhn, T. S. (2000). *The road since structure: Philosophical essays, 1970–1993, with an autobiographical interview*. Chicago: University of Chicago Press.

Langley, A. (1999). Strategies for theorizing from process data. *Academy of Management Review*, *24*, 691–711.

Lave, C. A., & March, J. G. (1975). *An introduction to models in the social sciences*. New York: Harper & Row.

Lee, P. (2004). *An introduction to Bayesian statistics*. New York: Hodder Arnold.

Levine, J. M., & Moreland, R. L. (2004). Collaboration: The social context of theory development. *Personality and Social Psychology Review*, *8*, 164–172.

Lewis, C. I. (1929). *Mind and the world order: Outline of a theory of knowledge*. New York: Dover.

Lewis, D. (2000). Causation as influence. *Journal of Philosophy*, *97*, 182–197.

Lewis, M., & Grimes, A. (1999). Meta-triangulation: Building theory from multiple paradigms. *Academy of Management Review*, *24*, 672–691.

Lickleiter, R., & Honeycutt, H. (2003). Developmental dynamics: Toward a biologically plausible evolutionary psychology. *Psychological Bulletin*, *129*, 819–835.

Lomi, A., & Larsen, E. (2001). *Dynamics of organizations: Computational modeling and organization theories*. Cambridge, MA: MIT Press.

Lubke, G. H., & Muthen, B. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods*, *10*, 21–39.

Luce, R. D., & Raiffa, H. (1957). *Games and decisions*. New York: Wiley.

Luke, S., Cioffi-Revilla, C., Panait, L., & Sullivan, K. (2004). *MASON: A new multi-agent simulation toolkit*. In *Proceedings of the 2004 Swarmfest Workshop* (pp. 181–187). Ann Arbor: University of Michigan.

MacCormack, C., & Strathern, C. (1980). *Nature, culture, and gender*. New York: Cambridge University Press.

MacCorquodale, K., & Meehl, P. (1948). On a distinction between hypothetical constructs and intervening variables. *Psychological Review*, *55*, 95–107.

Magidson, J., & Vermunt, J. (2007). *Introduction to latent class and finite mixture modeling*. New York: Chapman.

Mahfouz, A., Philaretou, A., & Theocharous, A. (2008). Virtual social interactions: Evolutionary, social psychological and technological perspectives. *Computers in Human Behavior*, *24*, 3014–3026.

Manicas, P. T. (2006). *A realist philosophy of science: Explanation and understanding*. Cambridge, UK: Cambridge University Press.

Marshall, D., & English, D. (2000). Neural network modeling of risk assessment in child protective services. *Psychological Methods*, *5*, 102–124.

Martin, R. M. (1979). *Studies in the scientific and mathematical philosophy of Charles S. Peirce: Essays by Carolyn Eisele*. The Hague, Netherlands: Mouton.

Martin, R. M., Hewstone, M., & Martin, P. (2003). Resistance to persuasive messages as a function of majority and minority source status. *Journal of Experimental Social Psychology*, *39*, 585–593.

Marx, K. (1887). *Das kapital* (Vol. 1). Moscow: Progress Publishers.

Marx, M. (1951). *Psychological theory: Contemporary readings*. New York: Macmillan.

Mason, B. (1996). Moving toward virtual ethnography. *American Folklore Society News*, *25*, 4–6.

Maxwell, J. (2004). Using qualitative methods for causal explanation. *Field Methods*, *16*, 243–264.

Mayo, C., & LaFrance, M. (1980). Toward an applicable social psychology. In R. F. Kidd & M. J. Saks (Eds.), *Advances in applied social psychology* (pp. 282–298). Hillsdale, NJ: Erlbaum.

McDonald, R. (1985). *Factor analysis and related methods*. Mahwah, NJ: Erlbaum.

McGrath, J. E., & Altman, I. (1966). *Small group research: A synthesis and critique of the field*. New York: Holt, Rinehart & Winston.

McGuire, W. J. (1968). Personality and susceptibility to social influence. In E. F. Borgatta & W. W. Mabert (Eds.), *Handbook of personality theory and research* (pp. 1130–1187). Chicago: Rand McNally.

McGuire, W. J. (1976). Some internal psychological factors influencing consumer choice. *Journal of Consumer Research*, *2*, 302–319.

McGuire, W. J. (1985). Attitudes and attitude change. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (pp. 233–346). New York: Random House.

McGuire, W. J. (1997). Creative hypothesis generating in psychology: Some useful heuristics. *Annual Review of Psychology*, *48*, 1–30.

McGuire, W. J. (2004). A perspectivist approach to theory construction. *Personality and Social Psychology Review*, *8*, 173–182.

McIntyre, A. (2007). *Participatory action research*. Thousand Oaks, CA: Sage.

Mead, G. (1932). *Mind, self and society from the standpoint of a social behaviorist*. Chicago: University of Chicago Press.

Mead, M. (2001). *Coming of age in Samoa: A psychological study of primitive youth for Western civilization*. New York: Harper.

Merton, R. (1968). *Social theory and social structure*. New York: Free Press.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81–97.

Millman, A., & Smith, C. (1997). Darwin's use of analogical reasoning in theory construction. *Metaphor and Symbol*, *12*, 159–187.

Mills, C. W. (1959). *The sociological imagination*. New York: Oxford University Press.

Miner, H. (1956). Body ritual among the Nacirema. *American Anthropologist*, *58*, 503–507.

Mintz, S. (1986). *Sweetness and power: The place of sugar in modern history*. New York: Penguin.

Mohr, L. (1982). *Explaining organizational behavior*. San Francisco: Jossey-Bass.

Monks, C. P., & Smith, P. (2006). Definitions of bullying: Age differences in understanding of the term and the role of experience. *British Journal of Developmental Psychology*, *24*, 801–821.

Morgan, S., & Winship, C. (2007). *Counterfactuals and causal inference*. New York: Cambridge University Press.

Morrin, M., Jacoby, J., Johar, G., He, X., Kuss, A., & Mazursky, D. (2002). Taking stock of stock brokers: Exploring investor information accessing strategies via process tracing. *Journal of Consumer Research*, *29*, 188–198.

Mueller, W. (2006). *Exploring precalculus. www.wmueller.com/precalculus/index.html.*

Murdock, G. P. (1954). *Outline of world cultures.* New York: Taplinger.

Murdock, G. P. (1957). World ethnographic sample. *American Anthropologist*, *59*, 4–28.

Murdock, G. P., & White, D. R. (1969). Standard cross-cultural sample. *Ethnology*, *8*, 329–369.

Muthen, B. O. (2001). Latent variable mixture modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 1–33). Mahwah, NJ: Erlbaum.

Muthen, B. O., & Muthen, L. K. (2000). Integrating person-centered and variable-centered analysis: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research*, *24*, 882–891.

Nakamura, J., & Csikszentmihalyi, M. (2001). Catalytic creativity: The case of Linus Pauling. *American Psychologist*, *56*, 337–341.

Nash, J. (1989). *From tank town to high tech: The clash of community and industrial cycles.* Albany: State University of New York Press.

Nersessian, N. (2002). Maxwell and "the method of physical analogy": Model-based reasoning, generic abstraction, and conceptual change. In D. Malament (Ed.), *Essays in the history and philosophy of science and mathematics* (pp. 129–166). La Salle, IL: Open Court.

Nersessian, N. (2008). *Creating scientific concepts.* Cambridge, MA: MIT Press.

Newell, A., & Simon, H. A., (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice-Hall.

Nisbett, R. E., & Wilson, T. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*, 231–259.

North, M. J., & Macal, C. (2005). Escaping the accidents of history: An overview of artificial life modeling with Repast. In A. Adamatzky & M. Komosinski (Eds.), *Artificial life models in software* (pp. 115–141). Heidelberg, Germany: Springer.

Nunnally, J. C. (1967). *Psychometric theory.* New York: McGraw-Hill.

Ortner, S. (1974). Is female to male as nature is to culture? In M. Rosado & L. Lamphere (Eds.), *Woman, culture and society* (pp. 68–87). Palo Alto, CA: Stanford University Press.

Osborn, A. (1963). *Applied imagination* (3rd ed.). New York: Scribner's.

Pap, A. (1962). *An introduction to the philosophy of science.* Glencoe, IL: Free Press of Glencoe (Macmillan).

Parsons, T. (1951). *The social system.* New York: Free Press.

Parsons, T. (1971). *The system of modern societies.* Englewood Cliffs, NJ: Prentice-Hall.

Patterson, G. (1975). *Families: Applications of social learning to family life.* Champaign, IL: Research Press.

Pearl, J. (2000). *Causality: Models, reasoning, and inference.* New York: Cambridge University Press.

Pearlson, G., & Folley, B. (2008). Schizophrenia, psychiatric genetics, and Darwinian psychiatry: An evolutionary framework. *Schizophrenia Bulletin*, *34*, 722–733.

Peter, J. P., & Olson, J. C. (1983). Is science marketing? *Journal of Marketing*, *47*, 111–125.

Peterson, C., & Seligman, M. E. P. (2004). *Character strengths and virtues: A classification and handbook.* New York: Oxford University Press/Washington, DC: American Psychological Association.

Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: The central and peripheral routes to attitude change.* New York: Springer-Verlag.

Petty, R. E., Cacioppo, J. T., & Goldman, R. (1981). Personal involvement as a determinant of argument based persuasion. *Journal of Personality and Social Psychology*, *41*, 847–855.

Platt, J. (1964). Strong inference. *Science*, *146*, 347–353.

Plattner, S. (1989). *Economic anthropology.* Palo Alto, CA: Stanford University Press.

Poincaré, H. (1952). *Science and hypothesis.* New York: Dover.

Poole, M., & Van de Ven, A. (1989). Using a paradox to build management and organization theories. *Academy of Management Review*, *14*, 562–579.

Popper, K. R. (1963). *Conjectures and refutations: The growth of scientific knowledge*. London: Routledge & Kegan Paul.

Popper, K. R. (1968). *The logic of scientific discovery*. London: Hutchinson.

Prochaska, J. O., DiClemente, C. C., & Norcross, J. C. (1992). In search of how people change: Applications to addictive behaviors. *American Psychologist*, *47*, 1102–1114.

Radecki, C. M., & Jaccard, J. (1995). Perceptions of knowledge, actual knowledge, and information search behavior. *Journal of Experimental Social Psychology*, *31*, 107–138.

Randall, W. (2007). From computer to compost: Rethinking our metaphors for memory. *Theory and Psychology*, *17*, 611–633.

Reason, P., & Bradbury, H. (2001). *Handbook of action research: Participative inquiry and practice*. Thousand Oaks, CA: Sage.

Reason, P., & Bradbury, H. (2007). *The Sage handbook of action research: Participative inquiry and practice*. Thousand Oaks: Sage.

Repenning, N. (2002). A simulation-based approach to understanding the dynamics of innovation implementation. *Organization Science*, *13*, 109–127.

Rivkin, J. W. (2000). Imitation of complex strategies. *Management Science*, *46*, 824–844.

Robins, G., Pattison, P., & Woodstock, J. (2005). Small and other worlds: Global network structures from local processes. *American Journal of Sociology*, *110*, 894–936.

Rogers, C. (1961). *On becoming a person*. New York: Houghton Mifflin.

Rollie, S., & Duck, S. (2006). Divorce and dissolution of romantic relationships: Stage models and their limitations. In M. Fine & J. Harvey (Eds.), *Handbook of divorce and relationship dissolution* (pp. 531–563). Mahwah, NJ: Erlbaum.

Root-Bernstein, R. S., & Root-Bernstein, M. (1999). *Sparks of genius: The thirteen thinking tools of the world's most creative people*. New York: Mariner.

Rosaldo, R. (1993). *Culture and truth: The remaking of social analysis*. New York: Beacon.

Rosenau, P. (1992). *Post-modernism and the social sciences*. Princeton, NJ: Princeton University Press.

Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*, 688–701.

Rubin, D. (1978). Bayesian inference for causal effects: The role of randomization, *Annals of Statistics*, *6*, 34–58.

Runkel, P. J., & McGrath, J. E. (1972). *Research on human behavior: A systematic guide to method*. New York: Holt, Rinehart & Winston.

Russell, B. (1931). *The scientific outlook*. London: Allen & Unwin.

Rychlak, J. (1994). *Logical learning theory: A human teleology and its empirical support*. Lincoln: University of Nebraska Press.

Sadovski, V. (1974). *Foundations of general systems theory*. Moscow: Nauka.

Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world*. Princeton, NJ: Princeton University Press.

Salmon, W. C. (1989). Four decades of scientific explanation. In P. Kitcher & W. C. Salmon (Eds.), *Scientific explanation* (pp. 3–196). Minneapolis: University of Minnesota Press.

Salmon, W. C. (1998). *Causality and explanation*. New York: Oxford University Press.

Salovey, P., & Grewal, D. (2005). The science of emotional intelligence. *Current Directions in Psychological Science*, *14*, 6–12.

Salovey, P., & Mayer, J. D. (1990). Emotional intelligence. *Imagination, Cognition, and Personality*, *9*, 185–211.

Sastry, M. A. (1997). Problems and paradoxes in a model of punctuated organizational change. *Administrative Science Quarterly*, *42*, 237–275.

Sayer, A. (1992). *Method in social science: A realist approach*. New York: Routledge.

Scarr, S. (1985). Constructing psychology: Making facts and fables for our times. *American Psychologist*, *40*, 499–512.

Schaefer, C., & DiGeron, T. (2000). *Ages and stages: A parent's guide to normal childhood development*. New York: Wiley.

Scheerer, M., Rothmann, E., & Goldstein, K. (1945). A case of "idiot-savant": An experimental study of personality organization. *Psychology Monographs*, *58*, 1–63.

Schiffman, S. S., Reynolds, M. L., & Young, F. W. (1981). *Introduction to multidimensional scaling*. New York: Academic Press.

Schimmack, U., & Crites, S. (2005). The structure of affect. In D. Albarracin, B. Johnson, & M. Zanna (Eds.), *The handbook of attitudes* (pp. 543–574). Mahwah, NJ: Erlbaum.

Schubert, G. (1960). *The public interest*. Glencoe, IL: Free Press.

Seligman, M. E. P. (2002). *Authentic happiness: Using the new positive psychology to realize your potential for lasting fulfillment*. New York: Free Press.

Seligman, M. E. P., & Csikszentmihalyi, M. (Eds.). (2000). Positive psychology [Special issue]. *American Psychologist*, *55*, 234–244.

Shadish, W. R., Cook, T., & Campbell, D. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Shaw, M., & Costanzo, P. (1982). *Theories of social psychology*. New York: McGraw-Hill.

Sheluga, D. A., & Jacoby, J. (1978). Do comparative claims encourage comparison shopping?: The impact of comparative claims on consumers' acquisition of product information. In J. Leigh & C. Martin (Eds.), *Current issues and research in advertising* (pp. 166–181). Ann Arbor: University of Michigan Press.

Sheth, J. N. (1967). A review of buyer behavior. *Management Science*, *13*, B718–B758.

Sierra, J., & Hyman, M. (2006). A dual-process model of cheating intentions. *Journal of Marketing Education*, *28*, 193–204.

Silverman, E., & Bryden, J. (2007). From artificial societies to new social science theory. *Advances in Artificial Life*, *22*, 565–574.

Simon, H. A. (1974). How big is a chunk? *Science*, *183*, 482–488.

Simon, H. A., & Newell, A. (1956) Models: Their uses and limitations. In L. D. White (Ed.), *The state of the social sciences* (pp. 61–83). Chicago: University of Chicago Press.

Simonton, D. K. (1988). *Scientific genius*. New York: Cambridge University Press.

Simonton, D. K. (2004). *Creativity in science: Chance, logic, genius, and zeitgeist*. New York: Cambridge University Press.

Skinner, B. F. (1957). *Verbal behavior*. Englewood Cliffs, NJ: Prentice-Hall.

Smith, E. R., & Conrey, F. (2007). Agent based modeling: A new approach for theory building in social psychology. *Personality and Social Psychology Review*, *11*, 87–104.

Smith, E. R., & DeCoster, J. (2000). Dual process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review, 4*, 103–131.

Smith, E. R., Zarate, M., & Branscombe, N. (1987). *Affect, ideology, and action: A dual-process model of attitude responses*. Technical Report, Purdue University, West Lafayette, IN.

Sorensen, I. (1998). *Thought experiments*. New York: Oxford University Press.

Sowa, J. (2000). *Processes and causality*. Retrieved April 2, 2006, from *www.jfsowa.com/ontology/causal.htm*.

Spatig, L., Parrott, L., Kusimo, P., Carter, C., & Keyes, M. (2001). We roll deep: Community and resistance in the lives of Appalachian girls. *Journal of Appalachian Studies*, *7*, 64–92.

Spiggle, S. (1994). Analysis and interpretation of qualitative data in consumer research. *Journal of Consumer Research*, *21*, 419–503.

Stein, M. (1974). *Stimulating creativity: I*. Oxford, UK: Academic Press.

Stein, M. (1975). *Stimulating creativity: II. Group procedures*. Oxford, UK: Academic Press.

Steinberg, L. (2008). A social neuroscience perspective on adolescent risk-taking. *Developmental Review, 28*, 78–106.

Stenstrom, E., Stenstrom, P., Saad, G., & Cheikhrouhou, S. (2008). Online hunting and gathering: An evolutionary perspective on sex differences in website preferences and navigation. *IEEE Transactions on Professional Communication*, *51*, 155–168.

Sterman, J. (2000). *Business dynamics: Systems thinking and modeling for a complex world*. New York: McGraw-Hill.

Sternberg, R., Grigorenko, E., & Singer, J. (2004). *Creativity: From potential to realization*. Washington, DC: American Psychological Association.

Sternberg, R. J. (1985). Implicit theories of intelligence, creativity, and wisdom. *Journal of Personality and Social Psychology*, *49*, 607–627.

Sternberg, R. J. (2002). Creativity as a decision. *American Psychologist*, *57*, 376.

Sternberg, R. J. (2003). *The psychologist's companion: A guide to scientific writing for students and researchers*. New York: Cambridge University Press.

Sternberg, R. J., & Lubart, T. (1996). Investing in creativity. *American Psychologist*, *51*, 677–688.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research*. Newbury Park, CA: Sage.

Sturdy, A., & Grey, C. (2003). Beneath and beyond organizational change management: Exploring alternatives. *Organization*, *10*, 651–662.

Suleiman, R., Troitzsch, K., & Gilbert, N. (2000). *Tools and techniques for social science simulation*. Heidelberg, Germany: Physica-Verlag.

Sullivan, P. F. (1991). On falsification interpretation of Peirce. *Transactions of the Charles S. Peirce Society*, *27*, 197–219.

Sun, R. (2006). *Cognition and multi-agent interaction: From cognitive modeling to social simulation*. New York: Cambridge University Press.

Takahashi, S., Kijima, K., & Sato, R. (2004). *Applied general systems research on organizations*. New York: Springer.

Tashakkori, A., & Teddlie, C. (2002). *Handbook of mixed methods: Social and behavioral research*. Thousand Oaks, CA: Sage.

Tetlock, P., & Belkin, A. (1996). *Counterfactual thought experiments in world politics*. Princeton, NJ: Princeton University Press.

Thom, R. (1975). *Structural stability and morphogenesis: An outline of a general theory of models*. Reading, MA: Benjamin.

Thomassen, L. (2006). *The Derrida–Habermas reader*. Chicago: University of Chicago Press.

Thurstone, L. L. (1947). *Multiple-factor analysis*. Chicago: University of Chicago Press.

Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: Wiley.

Toulmin, S. (1958). *The uses of argument*. Cambridge, UK: Cambridge University Press.

Turner, V. (1967). Carnival, ritual, and play in Rio de Janeiro. In A. Falassi (Ed.), *Time out of time: Essays on the festival* (pp. 74–92). Albuquerque: University of New Mexico Press.

Turner, V. (1970). *The forest of symbols: Aspects of Ndembu ritual*. Ithaca, NY: Cornell University Press.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*, 207–232.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131.

Underwood, B. (1957). *Psychological research*. New York: Appleton, Century, Crofts.

Van de Ven, A., & Poole, M. (1995). Explaining development and change in organizations *Academy of Management Review*, *20*, 510–541.

Van de Ven, A., & Poole, M. (2005). Alternative approaches for studying organizational change. *Organization Studies*, *26*, 1377–1399.

van Fraassen, B. C. (1980). *The scientific image*. Oxford, UK: Clarendon Press.

van Fraassen, B. C. (1989). *Laws and symmetry*. Oxford, UK: Clarendon Press.

Van Gigch, J. (1978). *Applied general systems theory*. New York: Harper & Row.

Veludo-de-Oliveira, T., Ikeda, A., & Campomar, M. (2006). Discussing laddering application by the means–end chain theory. *Qualitative Report*, *11*, 626–642.

Wadsworth, Y. (1998). *What is participatory action research?* Action Research International, Paper 2, New York.
Watzlawick, P. (1976). *How real is real?* New York: Random House.
Webb, E. J., Campbell, D., Schwartz, R., & Sechrest, L. (1966). *Unobtrusive measures: Nonreactive research in the social sciences*. Oxford, UK: Rand McNally.
Weber, M. (1904). *The Protestant ethic and the spirit of capitalism* (T. Parsons, Trans.). New York: Scribner's.
Weick, K. E. (1979). *The social psychology of organizing*. Reading, MA: Addison-Wesley.
Weisskopf, V. (1977). The frontiers and limits of science. *American Scientist*, *65*, 405–411.
West, C., & Zimmerman, D. (1987). Doing gender. *Gender and Society*, *1*, 125–151.
Westerhoff, F. (2008). The use of agent-based financial market models to test the effectiveness of regulatory policies. *Journal of Economics and Statistics, 228,* 195–227.
Wiggins, J. S. (1973). *Personality and prediction*. New York: Longman.
Wilson, T. D., Lisle, D. J., Schooler, J. W., Hodges, S. D., Klaaren, K. J., & LaFleur, S. J. (1993). Introspecting about reasons can reduce post-choice satisfaction. *Personality and Social Psychology Bulletin*, *19*, 331–339.
Wood, B. (2002). *Attachment and family systems: Family process*. New York: Wiley.
Wyer, R. S. (2004). A personalized theory of theory construction. *Personality and Social Psychology Review*, *8*, 201–209.
Zaheer, S., Albert, S., & Zaheer, A. (1999). Time scales and organizational theory. *Academy of Management Review*, *24*, 725–742.
Zaltman, G., & Coulter, R. (1995). Seeing the voice of the customer: Metaphor-based advertising research. *Journal of Advertising Research*, *83*, 35–51.
Zeeman, E. (1976). Catastrophe theory. *Scientific American*, *221*, 65–83.

Author Index

Subject Index

About the Authors

James Jaccard, PhD, is Professor of Psychology and Director of the Institute for Child Health and Development at Florida International University in Miami. Previously, he was Distinguished Professor of Psychology for 20 years at the State University of New York, Albany. Dr. Jaccard has authored or edited 11 books, including *Parental Monitoring of Adolescents*, *Dyadic Decision Making*, and *Interaction Effects in Multiple Regression*, and has published over 200 articles in peer-reviewed scientific journals. He has served on numerous boards and panels for the Institute of Medicine of the National Academies and the National Institutes of Health and is a Fellow of the American Psychological Association. His research focuses broadly on attitudes, cognitions, and emotions as they affect decision making, especially in applied settings. This includes research on adolescent decision making, health-related decisions, and a critical analysis of the effects of unconscious influences on adult decision making.

Jacob Jacoby, PhD, is Merchants Council Professor of Consumer Behavior at New York University's Stern School of Business. He has authored or edited six books, including *Brand Loyalty*, *The Comprehension and Miscomprehension of Print Communications*, and *Perceived Quality*, and has published over 160 articles in peer-reviewed social science and law journals. Dr. Jacoby is a Fellow of the American Psychological Association and a Fellow and past president of the Association for Consumer Research. His research on the factors that affect consumer decision making and behavior has been honored by awards from the American Psychological Association, the American Marketing Association, the American Academy of Advertising, the Association for Consumer Research, and the Society for Consumer Psychology. He has conducted research or consulted for dozens of Fortune 500 companies and other organizations (such as the American Association of Advertising Agencies and the National Football League) in the United States as well as major international companies abroad. He has also worked for federal agencies (including the U.S. Senate, Federal Trade Commission, and Food and Drug Administration) and testified in more than 100 cases heard in U.S. District Courts.