

[www.downloadslide.com](http://www.downloadslide.com)

Kenneth S. Bordens • Bruce B. Abbott



# Research Design and Methods

A Process Approach Eighth Edition



# Research Design and Methods

## *A Process Approach*

EIGHTH EDITION

Kenneth S. Bordens

Bruce B. Abbott

*Indiana University—Purdue University Fort Wayne*





RESEARCH DESIGN AND METHODS: A PROCESS APPROACH, EIGHTH EDITION

Published by McGraw-Hill, a business unit of The McGraw-Hill Companies, Inc., 1221 Avenue of the Americas, New York, NY 10020. Copyright © 2011 by The McGraw-Hill Companies, Inc. All rights reserved. Previous editions © 2008, 2005, 2002. No part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written consent of The McGraw-Hill Companies, Inc., including, but not limited to, in any network or other electronic storage or transmission, or broadcast for distance learning.

Some ancillaries, including electronic and print components, may not be available to customers outside the United States.

This book is printed on recycled, acid-free paper containing 10% postconsumer waste.

1 2 3 4 5 6 7 8 9 0 DOC/DOC 1 0 9 8 7 6 5 4 3 2 1 0

ISBN 978-0-07-353202-8

MHID 0-07-353202-9

Vice President & Editor-in-Chief: *Michael Ryan*

Vice President EDP/Central Publishing Services: *Kimberly Meriwether David*

Publisher: *Mike Sugarman*

Executive Editor: *Krista Bettino*

Managing Editor: *Meghan Campbell*

Executive Marketing Manager: *Pamela S. Cooper*

Senior Project Manager: *Lisa A. Brufodt*

Buyer: *Laura Fuller*

Design Coordinator: *Margarite Reynolds*

Media Project Manager: *Sridevi Palani*

Indexer: *Stephanie Abbott*

Compositor: *Laserwords Private Limited*

Typeface: *10/12 Goudy*

Printer: *R. R. Donnelley*

All credits appearing on page or at the end of the book are considered to be an extension of the copyright page.

Bordens, Kenneth S.

Research design and methods : a process approach / Kenneth S. Bordens, Bruce B. Abbott.—8th ed.  
p. cm.

Includes bibliographical references and index.

ISBN 978-0-07-353202-8 (alk. paper)

1. Psychology—Research—Textbooks. 2. Psychology—Research—Methodology—Textbooks.

I. Abbott, Bruce B. II. Title.

BF76.5.B67 2011

150.72—dc22

2010009326

*We dedicate this book to our parents, who provided us with the opportunity and inspiration to excel personally and professionally.*

*Lila Bordens and Walter Bordens  
Irene Abbott and Raymond Abbott*

*This page intentionally left blank*

# CONTENTS

---

*Preface* xvii

## **Chapter 1: Explaining Behavior 1**

What Is Science, and What Do Scientists Do?	2
Science as a Way of Thinking	3
How Do Scientists Do Science?	3
Basic and Applied Research	4
Framing a Problem in Scientific Terms	5
Learning About Research: Why Should You Care?	6
Exploring the Causes of Behavior	7
Explaining Behavior	8
Science, Nonscience, and Pseudoscience	9
Scientific Explanations	11
Commonsense Explanations Versus Scientific Explanations	14
Belief-Based Explanations Versus Scientific Explanations	16
When Scientific Explanations Fail	17
Failures Due to Faulty Inference	17
Pseudoexplanations	19
Methods of Inquiry	21
The Method of Authority	21
The Rational Method	21
The Scientific Method	22
The Scientific Method at Work: Talking on a Cell Phone and the Ability to Drive	25
The Steps of the Research Process	26
Summary	29
Key Terms	31

## **Chapter 2: Developing and Evaluating Theories of Behavior 32**

What Is a Theory?	32
Theory Versus Hypothesis	33
Theory Versus Law	34

Theory Versus Model	34
Mechanistic Explanations Versus Functional Explanations	37
Classifying Theories	38
Is the Theory Quantitative or Qualitative?	38
At What Level of Description Does the Theory Operate?	39
What Is the Theory's Domain?	43
Roles of Theory in Science	43
Understanding	43
Prediction	44
Organizing and Interpreting Research Results	44
Generating Research	44
Characteristics of a Good Theory	46
Ability to Account for Data	46
Explanatory Relevance	46
Testability	46
Prediction of Novel Events	47
Parsimony	47
Strategies for Testing Theories	48
Following a Confirmational Strategy	48
Following a Disconfirmational Strategy	49
Using Confirmational and Disconfirmational Strategies Together	49
Using Strong Inference	49
Theory-Driven Versus Data-Driven Research	51
Summary	54
Key Terms	55
<b>Chapter 3: Getting Ideas for Research</b>	<b>56</b>
Sources of Research Ideas	57
Experience	57
Theory	60
Applied Issues	62
Developing Good Research Questions	63
Asking Answerable Questions	64
Asking Important Questions	65
Developing Research Ideas: Reviewing the Literature	66
Reasons for Reviewing the Scientific Literature	66
Sources of Research Information	67
Performing Library Research	77
The Basic Strategy	77
Using PsycINFO	78
Using PsycARTICLES	80
Other Computerized Databases	80



General Internet Resources	81
Computer Searching for Books and Other Library Materials	81
Other Resources	82
Reading a Research Report	83
Obtaining a Copy	83
Reading the Literature Critically	84
Factors Affecting the Quality of a Source of Research Information	88
Publication Practices	89
Statistical Significance	89
Consistency With Previous Knowledge	91
Significance of the Contribution	92
Editorial Policy	93
Peer Review	93
Values Reflected in Research	96
Developing Hypotheses	98
Summary	99
Key Terms	101
<b>Chapter 4: Choosing a Research Design</b>	<b>102</b>
Functions of a Research Design	102
Causal Versus Correlational Relationships	103
Correlational Research	104
An Example of Correlational Research: Cell Phone Use and Motor Vehicle Accidents	105
Behavior Causation and the Correlational Approach	105
Why Use Correlational Research?	106
Experimental Research	108
Characteristics of Experimental Research	109
An Example of Experimental Research: Cell Phone Use While Driving	111
Strengths and Limitations of the Experimental Approach	112
Experiments Versus Demonstrations	113
Internal and External Validity	114
Internal Validity	114
External Validity	118
Internal Versus External Validity	119
Research Settings	120
The Laboratory Setting	120
The Field Setting	123
A Look Ahead	124
Summary	124
Key Terms	126

## Chapter 5: Making Systematic Observations 127

- Deciding What to Observe 127
- Choosing Specific Variables for Your Study 128
  - Research Tradition 128
  - Theory 128
  - Availability of New Techniques 129
  - Availability of Equipment 129
- Choosing Your Measures 130
  - Reliability of a Measure 130
  - Accuracy of a Measure 132
  - Validity of a Measure 133
  - Acceptance as an Established Measure 134
  - Scale of Measurement 135
  - Variables and Scales of Measurement 137
  - Choosing a Scale of Measurement 138
  - Adequacy of a Dependent Measure 141
  - Tailoring Your Measures to Your Research Participants 143
  - Types of Dependent Variables and How to Use Them 144
- Choosing When to Observe 148
- The Reactive Nature of Psychological Measurement 149
  - Reactivity in Research with Human Participants 149
  - Demand Characteristics 150
  - Other Influences 151
  - The Role of the Experimenter 152
  - Reactivity in Research with Animal Subjects 156
- Automating Your Experiments 157
- Detecting and Correcting Problems 158
  - Conducting a Pilot Study 158
- Adding Manipulation Checks 158
- Summary 159
- Key Terms 161

## Chapter 6: Choosing and Using Research Subjects 162

- General Considerations 162
  - Populations and Samples 163
  - Sampling and Generalization 164
  - Nonrandom Sampling 165
  - Is Random Sampling Always Necessary? 168
- Acquiring Human Participants for Research 168
  - The Research Setting 169
  - The Needs of Your Research 170
  - Institutional Policies and Ethical Guidelines 170

Voluntary Participation and Validity	171
Factors That Affect the Decision to Volunteer	171
Volunteerism and Internal Validity	174
Volunteerism and External Validity	176
Remedies for Volunteerism	177
Research Using Deception	178
Types of Research Deception	178
Problems Involved in Using Deception	179
Solutions to the Problem of Deception	181
Considerations When Using Animals as Subjects in Research	185
Contributions of Research Using Animal Subjects	186
Choosing Which Animal to Use	186
Why Use Animals?	187
How to Acquire Animals for Research	187
Generality of Animal Research Data	188
The Animal Rights Movement	189
Animal Research Issues	190
Alternatives to Animals in Research: In Vitro Methods and Computer Simulation	194
Summary	195
Key Terms	196

## **Chapter 7: Understanding Ethical Issues in the Research Process 197**

Ethical Research Practice With Human Participants	197
John Watson and Little Albert	197
Is It Fear or Is It Anger?	199
Putting Ethical Considerations in Context	199
The Evolution of Ethical Principles for Research With Human Participants	200
Nazi War Crimes and the Nuremberg Code	200
The Declaration of Helsinki	201
The Belmont Report	202
APA Ethical Guidelines	203
Government Regulations	203
Internet Research and Ethical Research Practice	207
Ethical Guidelines, Your Research, and the Institutional Review Board	210
Ethical Considerations When Using Animal Subjects	211
The Institutional Animal Care and Use Committee	212
Cost–Benefit Assessment: Should the Research Be Done?	213
Treating Science Ethically: The Importance of Research Integrity and the Problem of Research Fraud	214
What Constitutes Fraud in Research?	216
The Prevalence of Research Fraud	216

Explanations for Research Fraud 217

Dealing With Research Fraud 218

Summary 220

Key Terms 222

## **Chapter 8: Using Nonexperimental Research 223**

Conducting Observational Research 223

An Example of Observational Research: Are Children Really Cruel? 223

Developing Behavioral Categories 224

Quantifying Behavior in an Observational Study 225

Recording Single Events or Behavior Sequences 226

Coping With Complexity 226

Establishing the Reliability of Your Observations 229

Sources of Bias in Observational Research 234

Quantitative and Qualitative Approaches to Data Collection 235

Nonexperimental Research Designs 236

Naturalistic Observation 236

Ethnography 238

Sociometry 243

The Case History 244

Archival Research 245

Content Analysis 246

Meta-Analysis: A Tool for Comparing Results Across Studies 249

Step 1: Identifying Relevant Variables 251

Step 2: Locating Relevant Research to Review 252

Step 3: Conducting the Meta-Analysis 252

Drawbacks to Meta-Analysis 253

Summary 256

Key Terms 257

## **Chapter 9: Using Survey Research 258**

Survey Research 259

Designing Your Questionnaire 261

Writing Questionnaire Items 261

Assembling Your Questionnaire 267

Administering Your Questionnaire 269

Mail Surveys 269

Internet Surveys 270

Telephone Surveys 271

Group-Administered Surveys 271

Face-to-Face Interviews 272

A Final Note on Survey Techniques 273

Assessing the Reliability of Your Questionnaire	274
Assessing Reliability by Repeated Administration	274
Assessing Reliability With a Single Administration	275
Increasing Reliability	276
Assessing the Validity of Your Questionnaire	276
Acquiring a Sample for Your Survey	277
Representativeness	277
Sampling Techniques	278
Random and Nonrandom Sampling Revisited	285
Sample Size	286
Summary	287
Key Terms	289

## **Chapter 10: Using Between-Subjects and Within-Subjects Experimental Designs 290**

Types of Experimental Design	290
The Problem of Error Variance in Between-Subjects and Within-Subjects Designs	291
Sources of Error Variance	291
Handling Error Variance	293
Between-Subjects Designs	294
The Single-Factor Randomized-Groups Design	294
Matched-Groups Designs	299
Within-Subjects Designs	303
An Example of a Within-Subjects Design: Does Caffeine Keep Us Going?	303
Advantages and Disadvantages of the Within-Subjects Design	304
Sources of Carryover	306
Dealing With Carryover Effects	307
When to Use a Within-Subjects Design	313
Within-Subjects Versus Matched-Groups Designs	314
Types of Within-Subjects Designs	315
Factorial Designs: Designs With Two or More Independent Variables	317
An Example of a Factorial Design: Can That Witness Really Not Remember an Important Event?	318
Main Effects and Interactions	320
Factorial Within-Subjects Designs	322
Higher-Order Factorial Designs	323
Other Group-Based Designs	324
Designs With Two or More Dependent Variables	325
Confounding and Experimental Design	325

Summary 327

Key Terms 329

## **Chapter 11: Using Specialized Research Designs 330**

Combining Between-Subjects and Within-Subjects Designs 330

The Mixed Design 330

The Nested Design 332

Combining Experimental and Correlational Designs 335

Including a Covariate in Your Experimental Design 335

Including Quasi-Independent Variables in an Experiment 336

An Example of a Combined Design: Is Coffee a Physical or Psychological Stimulant? 336

Quasi-Experimental Designs 339

Time Series Designs 339

Equivalent Time Samples Design 340

Advantages and Disadvantages of Quasi Experiments 341

Nonequivalent Control Group Design 342

Pretest–Posttest Designs 343

Problems With the Pretest–Posttest Design 344

The Solomon Four-Group Design 346

Eliminating the Pretest 347

Developmental Designs 348

The Cross-Sectional Design 348

The Longitudinal Design 350

The Cohort-Sequential Design 353

Summary 354

Key Terms 356

## **Chapter 12: Using Single-Subject Designs 357**

A Little History 357

Baseline, Dynamic, and Discrete Trials Designs 359

Baseline Designs 360

An Example Baseline Experiment: Do Rats Prefer Signaled or Unsignaled Shocks? 361

Issues Surrounding the Use of Baseline Designs 363

Dealing With Uncontrolled Variability 366

Determining the Generality of Findings 368

Dealing With Problem Baselines 370

Types of Single-Subject Baseline Design 372

Dynamic Designs 380

Discrete Trials Designs 383

Characteristics of the Discrete Trials Design 383

Analysis of Data from Discrete Trials Designs 385

Inferential Statistics and Single-Subject Designs	386
Advantages and Disadvantages of the Single-Subject Approach	386
Summary	388
Key Terms	390

## Chapter 13: Describing Data 391

Descriptive Statistics and Exploratory Data Analysis	391
Organizing Your Data	392
Organizing Your Data for Computer Entry	396
Entering Your Data	398
Grouped Versus Individual Data	399
Graphing Your Data	400
Elements of a Graph	400
Bar Graphs	400
Line Graphs	402
Scatter Plots	404
Pie Graphs	404
The Importance of Graphing Data	405
The Frequency Distribution	406
Displaying Distributions	406
Examining Your Distribution	408
Descriptive Statistics: Measures of Center and Spread	410
Measures of Center	410
Measures of Spread	413
Boxplots and the Five-Number Summary	416
Measures of Association, Regression, and Related Topics	418
The Pearson Product-Moment Correlation Coefficient	418
The Point-Biserial Correlation	421
The Spearman Rank-Order Correlation	421
The Phi Coefficient	421
Linear Regression and Prediction	422
The Coefficient of Determination	424
The Correlation Matrix	425
Multivariate Correlational Techniques	426
Summary	426
Key Terms	428

## Chapter 14: Using Inferential Statistics 430

Inferential Statistics: Basic Concepts	430
Sampling Distribution	430
Sampling Error	431
Degrees of Freedom	431
Parametric Versus Nonparametric Statistics	431

The Logic Behind Inferential Statistics	432
Statistical Errors	434
Statistical Significance	435
One-Tailed Versus Two-Tailed Tests	436
Parametric Statistics	438
Assumptions Underlying a Parametric Statistic	438
Inferential Statistics With Two Samples	438
The $t$ Test	439
An Example from the Literature: Contrasting Two Groups	440
The $z$ Test for the Difference Between Two Proportions	441
Beyond Two Groups: Analysis of Variance (ANOVA)	442
The One-Factor Between-Subjects ANOVA	443
The One-Factor Within-Subjects ANOVA	447
The Two-Factor Between-Subjects ANOVA	448
The Two-Factor Within-Subjects ANOVA	451
Mixed Designs	451
Higher-Order and Special-Case ANOVAs	452
ANOVA Summing Up	452
Nonparametric Statistics	453
Chi-Square	453
The Mann–Whitney $U$ Test	455
The Wilcoxon Signed Ranks Test	456
Parametric Versus Nonparametric Statistics	456
Special Topics in Inferential Statistics	457
Power of a Statistical Test	457
Statistical Versus Practical Significance	459
The Meaning of the Level of Significance	459
Data Transformations	460
Alternatives to Inferential Statistics	461
Summary	464
Key Terms	465
<b>Chapter 15: Using Multivariate Design and Analysis</b>	<b>466</b>
Correlational and Experimental Multivariate Designs	466
Correlational Multivariate Design	467
Experimental Multivariate Design	468
Causal Inference	468
Assumptions and Requirements of Multivariate Statistics	469
Linearity	469
Outliers	469
Normality and Homoscedasticity	471
Multicollinearity	472
Error of Measurement	472
Sample Size	473



Correlational Multivariate Statistical Tests	474
Factor Analysis	474
Partial and Part Correlations	476
Multiple Regression	478
Discriminant Analysis	482
Canonical Correlation	483
Experimental Multivariate Statistical Tests	484
Multivariate Analysis of Variance	484
Multiway Frequency Analysis	489
Multivariate Statistical Techniques and Causal Modeling	491
Path Analysis	491
Structural Equation Modeling	495
Multivariate Analysis: A Cautionary Note	496
Summary	497
Key Terms	499

## Chapter 16: Reporting Your Research Results 500

APA Writing Style	500
Writing an APA-Style Research Report	501
Getting Ready to Write	501
Parts and Order of Manuscript Sections	503
The Title Page	504
The Abstract	505
The Introduction	506
The Method Section	510
The Results Section	513
The Discussion Section	517
The Reference Section	518
Footnotes	522
Tables	523
Figures	524
Elements of APA Style	526
Citing References in Your Report	526
Citing Quoted Material	527
Using Numbers in the Text	529
Avoiding Biased Language	530
Expression, Organization, and Style	531
Precision and Clarity of Expression	532
Economy of Expression	533
Organization	534
Style	535
Making It Work	536
Avoiding Plagiarism and Lazy Writing	538

Telling the World About Your Results	539
Publishing Your Results	539
Paper Presentations	540
The Ethics of Reporting or Publishing Your Results	542
Summary	542
Key Terms	544

*Appendix: Statistical Tables* A-1

*Glossary* G-1

*References* R-1

*Credits* C-1

*Name Index* I-1

*Subject Index* I-6

## PREFACE

---

This, the eighth edition of *Research Design and Methods: A Process Approach*, retains the general theme that characterized prior editions. As before, we take students through the research process, from getting and developing a research idea, to designing and conducting a study, through analyzing and reporting data. Our goals continue to be to present students with information on the research process in a lively and engaging way and to highlight the numerous decisions they must make when designing and conducting research. We also continue to stress how their early decisions in the process affect how data are collected, analyzed, and interpreted later in the research process. Additionally, we have continued the emphasis on the importance of ethical conduct, both in the treatment of research subjects and in the conduct of research and reporting research results.

In this edition we have retained the organization of topics, retaining the basic process approach. We have updated material in a number of chapters and updated many of the examples of research presented throughout the book. One change in the organization of the chapters is eliminating the list of questions that appeared at the end of each chapter in previous editions and salting them throughout each chapter. Students will find Questions to Ponder at various points in each chapter. These Questions to Ponder have students reflect on the material they read in the preceding section and allow students to prepare themselves for the material to follow. We believe that redistributing the questions in this way will help students better understand the material they read.

### CHANGES IN THE EIGHTH EDITION

---

We have revised each chapter by updating examples and revising material where appropriate, as described below.

### CHAPTER 1: EXPLAINING BEHAVIOR

---

A new introductory vignette focusing on the timely issue of texting while driving opens the chapter and is carried through the chapter where appropriate. We have rewritten the section on explaining behavior. This section now opens with an example (EMDR therapy) to get students thinking about how science is applied to explain behavior. The EMDR example is then used to illustrate the differences between real science and pseudoscience and how scientific explanations are developed.

## CHAPTER 2: DEVELOPING AND EVALUATING THEORIES OF BEHAVIOR

---

A more recent example of a proposed scientific law (Herrnstein's "matching law") has been substituted for Thorndike's "law of effect," and recent applications of the matching law in basketball and football are described. In the section describing the characteristics of a good theory, the example of the ability of a theory to predict novel events has been changed from Einstein's theory of relativity to the Rescorla-Wagner model of classical conditioning, in which the model's counterintuitive prediction of "overexpectation" was confirmed.

## CHAPTER 3: GETTING IDEAS FOR RESEARCH

---

This chapter remains largely unchanged from the previous edition. We have updated the section on using PsycINFO. In this section we eliminated the example of a PsycINFO entry in order to tighten the chapter. We have also updated the section on the peer review process by including a reference to a 2009 paper by Suls and Martin on the problems of the traditional peer review process.

## CHAPTER 4: CHOOSING A RESEARCH DESIGN

---

The topic of the dangers of cell-phone use while driving is carried over from the opening vignette of Chapter 1 with a pair of new examples: The correlational approach is illustrated by research on the incidence of motor vehicle accidents resulting in substantial damage (Redelmeier & Tibshirani, 1997) or hospital attendance (McEvoy, Stevenson, McCartt, et al., 2005) at or near the time that the driver's cell phone was in use as indicated by phone-company records. The experimental approach is illustrated research using a highly realistic driving simulator to test driver reactions while conversing with a friend either via cell phone or with the friend as passenger (Strayer & Drews, 2007).

## CHAPTER 5: MAKING SYSTEMATIC OBSERVATIONS

---

This chapter is unchanged from the seventh edition except for minor improvements in wording.

## CHAPTER 6: CHOOSING AND USING RESEARCH SUBJECTS

---

Chapter 6 continues to focus on issues relating to using subjects/participants in the research process (e.g., sampling, volunteer bias, research deception, and using animals in research). We have updated the section on volunteer bias by including references to recent research on the impact of volunteerism in various types of research. Similarly, the section on using deception in research has been updated to include new references on the problem of deception and how to reduce the impact of deception. The section on the animal rights issue has also been updated.

## **CHAPTER 7: UNDERSTANDING ETHICAL ISSUES IN THE RESEARCH PROCESS**

---

The material on the history of ethical issues has been condensed. We eliminated the extended table summarizing government regulations on using human research participants (but provided a link to the HHS Web site for interested students). The section on Institutional Review Boards has been updated by adding a reference to a 2009 article showing how the IRB benefits researchers.

## **CHAPTER 8: USING NONEXPERIMENTAL RESEARCH**

---

The section on content analysis has been updated to reflect the emergence of popular Internet resources such as blogs and social networking sites in addition to Web pages as important sources of material for content analysis.

## **CHAPTER 9: USING SURVEY RESEARCH**

---

A new example opens the chapter. The new example focuses on a survey of how Americans obtained political information leading up to the 2008 presidential election. This new example is then used throughout the chapter. The section on Internet surveys has been updated to include an expanded discussion of the differences and similarities between the results from traditional and Internet survey methods.

## **CHAPTER 10: USING BETWEEN-SUBJECTS AND WITHIN-SUBJECTS EXPERIMENTAL DESIGNS**

---

This chapter has been updated with fresh and entertaining examples of the multiple control group design (Balcetis & Dunning, 2007), the factorial between-subjects design (Kassam, Gilbert, Swencionis, & Wilson, 2009), and the factorial within-subjects design (Berman, Jonides, and Kaplan, 2008).

## **CHAPTER 11: USING SPECIALIZED RESEARCH DESIGNS**

---

A number of figures illustrating various time-series designs have been redone to improve clarity.

## **CHAPTER 12: USING SINGLE-SUBJECT DESIGNS**

---

A study by Hoch and Taylor (2008) has been added as an example of the use of an ABAB design in an applied setting and integrated into the discussion. (The study evaluated a technique for getting teenagers with autism to eat their meals at a normal rate rather than

wolfing the meals down.) A new section has been added on judging stable differences in performance across phases, citing concerns about the ability of researchers to judge differences in baseline levels across treatments accurately, and describing suggested solutions. The section on inferential statistics and single-subject designs has been updated to reflect current opinion on this topic.

## **CHAPTER 13: DESCRIBING DATA**

---

The discussions of bar graphs and line graphs have been revised to reflect the recent emphasis on including some measure of precision in these graphs. The section on scatter plots was expanded slightly to describe the possible inclusion of a regression line on the graph.

## **CHAPTER 14: USING INFERENTIAL STATISTICS**

---

The section on effect size has been expanded slightly to highlight the recent strong recommendation by the American Psychological Association to include measures of effect size wherever possible and appropriate. Some discussions have been slightly rewritten to improve clarity.

## **CHAPTER 15: USING MULTIVARIATE DESIGN AND ANALYSIS**

---

The section on multivariate statistical tests for experimental designs now includes multiway frequency analysis. Structural equation modeling is now mentioned along with path analysis and a description of its use.

## **CHAPTER 16: REPORTING YOUR RESEARCH RESULTS**

---

This chapter has been significantly revised to reflect the changes in the sixth edition of the publication manual of the American Psychological Association. A new research example is used for the sample paper appearing in the relevant figures illustrating the various sections of an APA-style paper.

## **ANCILLARIES**

---

The ancillaries continue to be provided via the McGraw-Hill Web site at [www.mhhe.com/bordens8e](http://www.mhhe.com/bordens8e). Students will have access to an updated study guide reflecting the changes made to the content and organization of the text. Each chapter of the guide includes a list of key terms, practice questions (multiple-choice, fill-in, and essay), and hands-on exercises. Instructors will have access to an instructor's manual, test bank, and PowerPoint presentations, all developed by the authors. These have all been updated to reflect the changes made to the text.

## ACKNOWLEDGMENTS

---

After eight editions the list is long of those to whom we owe our thanks—past reviewers, editors, colleagues, and students who have contributed their time and talents to improve the text and make it successful. For the eighth edition we especially wish to single out the reviewers: Elizabeth Arnott-Hill, Chicago State University; Nicole Avena, North Central University; Scott Bates, Utah State University, Logan; Garrett Berman, Roger Williams University; Elliot Bonem, Michigan State University; Amy M. Buddie, Kennesaw State University; Anastasia Dimitriopoulos, Case Western Reserve University; William Dragon, Cornell College; Richard M. Flicker, Southern University–Baton Rouge; Harvey Ginsburg, Southwest Texas State University; Michael Hall, James Madison University; Gregory Hundt, High Point University; Michael Jarvinen, University of Michigan, Flint; Derek Mace, Kutztown University; Bradley D. McAuliff, California State University, Northridge; Ryan Newell, Oklahoma Christian University of Science and Arts; Carlota Ocampo, Trinity University; Susan Parault, St. Cloud State University; Kerri Pickel, Ball State University; Judith Platania, Roger Williams University; Christopher Poirier, Stonehill College; Christine Selby, Husson College; Royce Simpson, Spring Hill College; Shannon Whitten, Brevard Community College–Palm Bay Campus; Josephine Wilson, Wittenberg University; William Wozniak, University of Nebraska, Kearny; Minhnoi Wroble-Biglan, Pennsylvania State University–Beaver Campus; Lorienta Yancura, University of Hawaii, Manoa; Karen Yanowitz, Arkansas State University. Their criticisms and suggestions have been greatly appreciated. Our thanks go also to Mike Sugarman, Publisher at McGraw-Hill; to Stephanie Kelly, Development Editor, Triple SSS Press Media Development, Inc.; to our copy editor, Alyson Platt, who worked tirelessly to correct and improve our manuscript; and to those other members of the McGraw-Hill staff who worked on the ancillaries and organized the text Web site where these resources are made available to students and instructors.

Finally, we offer a special thanks to our wives, Stephanie Abbott and Ricky Karen Bordens, for their support and encouragement, and to our families.

**Kenneth S. Bordens**  
**Bruce B. Abbott**

*This page intentionally left blank*



# Explaining Behavior

The night of June 26, 2007, was supposed to be one of celebration for Bailey Goodman and her four friends. After all, she and her friends were driving to her parents' lake cottage to celebrate their graduation from Fairport High School near Rochester, New York. Their plans were to spend a few days together at the cottage and then return home to attend some graduation parties. The future looked bright for the five young women, all of whom were cheerleaders at their high school. Unfortunately, those bright futures were not to be realized. On their way to the cottage, Bailey Goodman, who was driving an SUV, crossed over the centerline of the road and crashed head-on into an oncoming tractor trailer truck driven by 50-year-old David Laverty. Moments after the catastrophic collision, Goodman's SUV burst into flames, trapping the girls in the burning wreckage. All five were killed in the inferno.

Truck driver Laverty saw the oncoming SUV in the distance pass another vehicle, making it safely back to its own lane. He thought little more of the oncoming SUV until it veered suddenly into his lane. It happened so fast that Laverty had no time to react. An investigation into the crash by the local sheriff ruled out Laverty as a cause of the accident. Autopsies showed that Goodman was not drunk nor was she impaired by drugs. However, the investigation did turn up a possible explanation for why Goodman veered into the truck's path. When Goodman's cell phone records were reviewed, investigators discovered that Goodman had sent a text message at 10:05 P.M. and that she had received a reply at 10:06 P.M. The first report of the crash, made by another friend of Goodman who was following in another vehicle, came in at 10:07 P.M. Investigators believed that Goodman was "driving while texting." Goodman may have been distracted by the text and failed to notice that her vehicle was drifting over the centerline. Of course, investigators had no way of determining if Goodman was the one actually using the phone at the time of the crash. The sequence of events, however, provides a plausible explanation for the accident.

## CHAPTER OUTLINE

### What Is Science, and What Do Scientists Do?

- Science as a Way of Thinking
- How Do Scientists Do Science?
- Basic and Applied Research
- Framing a Problem in Scientific Terms

### Learning About Research: Why Should You Care?

#### Exploring the Causes of Behavior

#### Explaining Behavior

- Science, Nonscience, and Pseudoscience
- Scientific Explanations
- Commonsense Explanations Versus Scientific Explanations
- Belief-Based Explanations Versus Scientific Explanations

#### When Scientific Explanations Fail

- Failures Due to Faulty Inference
- Pseudoexplanations

#### Methods of Inquiry

- The Method of Authority
- The Rational Method
- The Scientific Method
- The Scientific Method at Work: Talking on a Cell Phone and the Ability to Drive
- The Steps of the Research Process

#### Summary

#### Key Terms

The sad fate of Bailey Goodman and her friends is not unique. There are numerous other examples of accidents resulting from people talking on a cell phone or texting while driving. In fact, many states have or are considering laws restricting cell phone use while driving.

The issue of using a cell phone while driving raises a question about the human being's capacity to "multitask"—do more than one thing at a time. Based on the Bailey Goodman story and others like it, we could engage in endless speculation about whether such multitasking is a general problem for everyone or unique to those who are hurt or killed in the attempt. Was Goodman's relative inexperience as a new driver a major factor in the accident? Would a more experienced driver be able to handle the multitasking better than Goodman? Although such speculations make for interesting dinner table conversation, they do nothing to address the basic question concerning distraction while multitasking and how it relates to a driver's ability to drive a car.

Questions such as the one about one's ability to multitask (talk on the phone while driving) almost cry out for answers. This is where science and scientists come in. When confronted with situations such as Bailey Goodman's, scientists are curious. Like most of us, they wonder if there is a relationship between the distraction of talking or texting on a cell phone and driving ability. Scientists, however, go beyond mere speculation: they formulate ways to determine clearly the relationship between talking on a cell phone and driving ability and then design research studies to test the relationship.

This book is about how the initial curiosity sparked by an event such as the Goodman accident gets transformed into a testable research question and eventually into a research study yielding data that are analyzed. Only through this process can we move beyond dinner table speculations and into the realm of scientific explanation.

## WHAT IS SCIENCE, AND WHAT DO SCIENTISTS DO?

---

The terms *science* and *scientist* probably conjure up a variety of images in your mind. A common image is that of a person in a white lab coat surrounded by bubbling flasks and test tubes, working diligently to discover a cure for some dreaded disease. Alternatively, our lab-coated scientist might be involved in some evil endeavor that will threaten humankind. Books, movies, and television have provided such images. Just think about the classic horror films of the 1940s and 1950s (e.g., *Frankenstein*), and it is not hard to see where some of these images come from.

Although these images may be entertaining, they do not accurately capture what science actually is and what real scientists do. Simply put, **science** is a set of methods used to collect information about phenomena in a particular area of interest and build a reliable base of knowledge about them. This knowledge is acquired via *research*, which involves a scientist identifying a phenomenon to study, developing hypotheses, conducting a study to collect data, analyzing the data, and disseminating the results. Science also involves developing theories to help better describe, explain, and organize scientific information that is collected. At the heart of any science (psychology included) is information that is obtained through observation and measurement

of phenomena. So, for example, if I want to know if text messaging while driving is a serious threat to safety, I must go out and make relevant observations. Science also requires that any explanations for phenomena can be modified and corrected if new information becomes available. Nothing in science is taken as an absolute truth. All scientific observations, conclusions, and theories are always open to modification and perhaps even abandonment as new evidence arises.

Of course, a **scientist** is someone who does science. A scientist is a person who adopts the methods of science in his or her quest for knowledge. However, this simple definition does not capture what scientists do. Despite the stereotyped image of the scientist hunkered over bubbling flasks, scientists engage in a wide range of activities designed to acquire knowledge in their fields. These activities take place in a variety of settings and for a variety of reasons. For example, you have scientists who work for pharmaceutical companies trying to discover new medications for the diseases that afflict humans. You have scientists who brave the bitter cold of the Arctic to take ice samples that they can use to track the course of global climate change. You have scientists who sit in observatories with their telescopes pointed to the heavens, searching for and classifying celestial bodies. You have scientists who work at universities and do science to acquire knowledge in their chosen fields (e.g., psychology, biology, or physics). In short, science is a diverse activity involving a diverse group of people doing a wide range of things. Despite these differences, all scientists have a common goal: to acquire knowledge through the application of scientific methods and techniques.

### Science as a Way of Thinking

It is important for you to understand that science is not just a means of acquiring knowledge; it is also a way of thinking and of viewing the world. A scientist approaches a problem by carefully defining its parameters, seeking out relevant information, and subjecting proposed solutions to rigorous testing. The scientific view of the world leads a person to be skeptical about what he or she reads or hears in the popular media. Having a scientific outlook leads a person to question the validity of provocative statements made in the media and to find out what scientific studies say about those statements. In short, an individual with a scientific outlook does not accept everything at face value.

The scientific method is not the only way to approach a problem. As we discuss later in this chapter, some problems (philosophical, ethical, or religious) may not lend themselves to exploration with the scientific method. In those cases, other methods of inquiry may be more useful.

### How Do Scientists Do Science?

In their quest for knowledge about a phenomenon, scientists can use a wide variety of techniques, each suited to a particular purpose. Take the question about using a cell phone while driving an automobile. You, as a scientist, could approach this issue in several ways. For example, you could examine public records on automobile accidents and record the number of times a cell phone was in use at the time of the accident. You would then examine your data to see if there is a relationship between talking on

a cell phone and having an automobile accident. If you found that there was a greater frequency of accidents when drivers were talking on a cell phone, this would verify the role of cell phones in automobile accidents.

Another way you could approach this problem is to conduct a controlled experiment. You could have participants perform a simulated driving task and have some drivers talk on a cell phone and others not. You could record the number of driving errors made. If you found a greater number of errors on the driving task when the drivers were talking on the cell phone, you would have verified the effect on driving ability of talking on a cell phone.

## QUESTIONS TO PONDER

---

1. What is science, and what do scientists do?
2. What is meant by the statement that the scientific method is an attitude? (Explain)
3. How do scientists obtain knowledge on issues that interest them?

## Basic and Applied Research

Scientists work in a variety of areas to identify phenomena and develop valid explanations for them. The goals established by scientists working within a given field of research may vary according to the nature of the research problem being considered. For example, the goal of some scientists is to discover general laws that explain particular classes of behaviors. In the course of developing those laws, psychologists study behavior in specific situations and attempt to isolate the variables affecting behavior. Other scientists within the field are more interested in tackling practical problems than in finding general laws. For example, they might be interested in determining which of several therapy techniques is best for treating severe phobias.

An important distinction has been made between basic research and applied research along the lines just presented.

**Basic Research** **Basic research** is conducted to investigate issues relevant to the confirmation or disconfirmation of theoretical or empirical positions. The major goal of basic research is to acquire general information about a phenomenon, with little emphasis placed on applications to real-world examples of the phenomenon (Yaremko, Harari, Harrison, & Lynn, 1982). For example, research on the memory process may be conducted to test the efficacy of interference as a viable theory of forgetting. The researcher would be interested in discovering something about the forgetting process while testing the validity of a theoretical position. Applying the results to forgetting in a real-world situation would be of less immediate interest.

**Applied Research** The focus of **applied research** is somewhat different from that of basic research. Although you may still work from a theory when formulating your hypotheses, your primary goal is to generate information that can be applied directly

to a real-world problem. A study by James Ogloff and Neil Vidmar (1994) on pretrial publicity provides a nice example of applied research. It informs us about a very real problem facing the court system: To what extent does pretrial publicity affect the decisions jurors make about a case? The results of studies such as Ogloff and Vidmar's can help trial and appeals court judges make decisions concerning limitations placed on jury exposure to pretrial publicity. Further examples of applied research can be found in the areas of clinical, environmental, and industrial psychology (among others).

**Overlap Between Basic and Applied Research** The distinction between applied and basic research is not always clear. Some research areas have both basic and applied aspects. Consider the work of Elizabeth Loftus (1979) on the psychology of the eyewitness. Loftus has extensively studied the factors that affect the ability of an eyewitness to accurately perceive, remember, and recall a criminal event. Her research certainly fits the mold of applied research. But her results also have some implications for theories of memory, so they also fit the mold of basic research. In fact, many of Loftus's findings can be organized within existing theories of memory.

Even applied research is not independent of theories and other research in psychology. The defining quality of applied research is that the researcher attempts to conduct a study the results of which can be applied directly to a real-world event. To accomplish this task, you must choose a research strategy that maximizes the applicability of findings.

### Framing a Problem in Scientific Terms

Kelly (1963) characterizes each person as a scientist who develops a set of strategies for determining the causes of behavior observed. We humans are curious about our world and like to have explanations for the things that happen to us and others. After reading about Bailey Goodman's accident, you may have thought about potential explanations for the accident. For example, you might have questioned Goodman's competence as a driver or speculated about the role of alcohol or drugs in the accident.

Usually, the explanations we come up with are based on little information and mainly reflect personal opinions and biases. The everyday strategies we use to explain what we observe frequently lack the rigor to qualify as truly scientific approaches. In most cases, the explanations for everyday events are made on the spot, with little attention given to ensuring their accuracy. We simply develop an explanation and, satisfied with its plausibility, adopt it as true. We do not consider exploring whether our explanation is correct or whether there might be other, better explanations.

If we do give more thought to our explanations, we often base our thinking on hearsay, conjecture, anecdotal evidence, or unverified sources of information. These revised explanations, even though they reduce transient curiosity, remain untested and are thus of questionable validity. In the Bailey Goodman case you might conclude that talking on a cell phone while driving distracts the driver from important tasks required to successfully navigate a car. Although this explanation seems plausible, without careful testing it remains mere speculation. To make matters worse, we have a tendency to look for information that will confirm our prior beliefs and

assumptions and to ignore or downplay information that does not conform to those beliefs and assumptions. So, if you believe that talking on cell phones causes automobile accidents, you might seek out newspaper articles that report on such accidents and fail to investigate the extent to which cell phone use while driving does not lead to an accident. The human tendency to seek out information that confirms what is already believed is known as **confirmation bias**. At the same time, you may ignore information that conflicts with your beliefs.

Unfounded but commonly accepted explanations for behavior can have widespread consequences when the explanations become the basis for social policy. For example, segregation of Blacks in the South was based on stereotypes of assumed racial differences in intelligence and moral judgment. These beliefs sound ludicrous today and have failed to survive a scientific analysis. Such mistakes might have been avoided if lawmakers of the time had relied on objective information rather than on prejudice.

To avoid the trap of easy, untested explanations for behavior, we need to abandon the informal, unsystematic approach to explanation and adopt an approach that has proven its ability to find explanations of great power and generality. This approach, called the *scientific method*, and how you can apply it to answer questions about behavior are the central topics of this book.

## LEARNING ABOUT RESEARCH: WHY SHOULD YOU CARE?

---

Students sometimes express the sentiment that learning about research is a waste of time because they do not plan on a career in science. Although it is true that a strong background in science is essential if you plan to further your career in psychology after you graduate, it is also true that knowing about science is important even if you do *not* plan to become a researcher.

The layperson is bombarded by science every day. When you read about the controversy over stem-cell research or global warming, you are being exposed to science. When you read about a “scientific” poll on a political issue, you are being exposed to science. When you hear about a new cure for a disease, you are being exposed to science. When you are persuaded to buy one product over another, you are being exposed to science. Science, on one level or another, permeates our everyday lives. To deal rationally with your world, you must be able to analyze critically the information thrown at you and separate scientifically verified facts from unverified conjecture.

Often, popular media such as television news programs present segments that *appear* scientific but on further scrutiny turn out to be flawed. One example was a segment on the ABC television news show *20/20* on sexual functions in women after a hysterectomy. In the segment, three women discussed their posthysterectomy sexual dysfunction. One woman reported, “It got to the point where I couldn’t have sex. I mean, it was so painful . . . we couldn’t do it.” The testimonials of the three patients were backed up by a number of medical experts who discussed the link between hysterectomy and sexual dysfunction.

Had you watched this segment and looked no further, you would have come away with the impression that posthysterectomy sexual dysfunction is common. After all, all the women interviewed experienced it, and the experts supported them.

However, your impression would not be correct. When we examine the research on posthysterectomy sexual functioning, the picture is not nearly as clear as the one portrayed in the *20/20* segment. In fact, there are studies showing that after hysterectomy, women may report an *improvement* in sexual function (Rhodes, Kjerulff, Langenberg, & Guzinski, 1999). Other studies show that the type of hysterectomy a woman has undergone makes a difference. If the surgery involves removing the cervix (a total hysterectomy), there is more sexual dysfunction after surgery than if the cervix is left intact (Saini, Kuczynski, Gretz, & Sills, 2002). Finally, the Boston University School of Medicine's Institute for Sexual Medicine reports that of 1,200 women seen at its Center for Sexual Medicine, very few of them complained of posthysterectomy sexual dysfunction (Goldstein, 2003).

As this examples suggests, whether you plan a career in research or not, it is to your benefit to learn how research is done. This will put you in a position to evaluate information that you encounter that is supposedly based on "science."

## EXPLORING THE CAUSES OF BEHAVIOR

---

Psychology is the science of behavior and mental processes. The major goals of psychology (as in any other science) are (1) to build an organized body of knowledge about its subject matter and (2) to develop valid, reliable explanations for the phenomena within its domain. For example, psychologists interested in aggression and the media would build a storehouse of knowledge concerning how various types of media violence (e.g., movies, television shows, cartoons, or violent video games) affect aggressive behavior. If it were shown that violent video games do increase aggression, the psychologist would seek to explain how this occurs.

How do you, as a scientist, go about adding to this storehouse of knowledge? The principal method for acquiring knowledge and uncovering causes of behavior is **research**. You identify a problem and then systematically set out to collect information about the problem and develop explanations.

Robert Cialdini (1994) offers a simple yet effective analogy to describe the process of studying behavior: He likens science to a hunting trip. Before you go out to "bag" your prey, you must first scout out the area within which you are going to hunt. On a hunting trip, scouting involves determining the type and number of prey available in an area. Cialdini suggests that in science "scouting" involves making systematic observations of naturally occurring behavior.

Sometimes scouting may not be necessary. Sometimes the prey falls right into your lap without you having to go out and find it. Cialdini tells a story of a young woman who was soliciting for a charity. Initially, Cialdini declined to give a donation. However, after the young woman told him that "even a penny would help," he found himself digging into his wallet. As he reflected on this experience, he got to wondering why he gave a donation after the "even a penny would help" statement. This led him to a series of studies on the dynamics of compliance. In a similar manner, as you read about the Bailey Goodman case, you might already have begun to wonder about the factors that contribute to distraction-related automobile accidents. As we describe in Chapter 3, "scouting" can involve considering many sources.

The second step that Cialdini identifies is “trapping.” After you have identified a problem that interests you, the next thing to do is identify the factors that might affect the behavior that you have scouted. Then, much like a hunter closing in on prey, you systematically study the phenomenon and identify the factors that are crucial to explaining that phenomenon. For example, after wondering whether talking on a cell phone causes automobile accidents you could set up an experiment to test this. You could have participants do a simulated driving task. Participants in one condition would do the simulated driving task while talking on a cell phone, and participants in another would do the task without talking on a cell phone. You could record the number of errors participants make on the simulated driving task. If you find that participants talking on a cell phone make more errors than those not talking on a cell phone, you have evidence that talking on a cell phone while driving causes drivers to make more potentially dangerous errors.

### QUESTIONS TO PONDER

---

1. How do basic and applied research differ, and how are they similar?
2. How are problems framed in research terms?
3. What is confirmation bias, and what are its implications for understanding behavior?
4. Why should you care about learning about research, even if you are not planning a career in research?
5. What are the two steps suggested by Cialdini (1994) for exploring the causes of behavior, and how do they relate to explaining behavior?

### EXPLAINING BEHAVIOR

---

Imagine that after being in an automobile accident (perhaps caused by your friend who was texting while driving) you find yourself depressed, unable to sleep, and lacking appetite. After a few weeks of feeling miserable, you find a therapist whom you have heard can help alleviate your symptoms. On the day of your appointment you meet with your new therapist. You begin by mapping out a therapy plan with your therapist. You and she identify stressful events you have experienced, present situations that are distressing to you, and events in your past that might relate to your current symptoms. Next you identify an incident that is causing you the most distress (in this case your accident) and your therapist has you visualize things relating to your memory of the event. She also has you try to reexperience the sensations and emotions related to the accident.

So far you are pretty satisfied with your therapy session because your therapist is using techniques you have read about and that are successful in relieving symptoms like yours. What occurs next, however, puzzles you. Your therapist has you follow her finger with your eyes as she moves it rapidly back and forth across your field of



vision. Suddenly, she stops and tells you to let your mind go blank and attend to any thoughts, feelings, or sensations that come to mind. You are starting to wonder just what is going on. Whatever you come up with, your therapist tells you to visualize and has you follow her finger once again with your eyes. On your way home after the session you wonder just what the finger exercise was all about.

When you get home, you do some research on the Internet and find that your therapist was using a technique called Eye Movement Desensitization and Reprocessing (EMDR) therapy. You read that the eye movements are supposed to reduce the patient's symptoms rapidly. Because you did not experience this, you decide to look into what is known about EMDR therapy. What you find surprises you. You find a number of Web sites touting the effectiveness of EMDR. You read testimonials from therapists and patients claiming major successes using the treatment. You also learn that many clinical psychologists doubt that the eye movements are a necessary component of the therapy. In response, advocates of EMDR have challenged critics to prove that EMDR does not work. They suggest that those testing EMDR are not properly trained in the technique, so it will not work for them. They also suggest that the eye movements are not necessary and that other forms of stimulation, such as the therapist tapping her fingers on the client's leg, will work. You are becoming skeptical. What you want to find is some real scientific evidence concerning EMDR.

### Science, Nonscience, and Pseudoscience

We have noted that one goal of science is to develop explanations for behavior. This goal is shared by other disciplines as well. For example, historians may attempt to explain why Robert E. Lee ordered Pickett's Charge on the final day of the Battle of Gettysburg. Any explanation would be based on reading and interpreting historical documents and records. However, unless such explanations can be submitted to empirical testing, they are not considered scientific.

What distinguishes a true science from nonscience and pseudoscience? The difference lies in the methods used to collect information and draw conclusions from it. A true science (such as psychology, physics, chemistry, and biology) relies on established scientific methods to acquire information and adheres to certain rules when determining the validity of information acquired.

A *nonscience* can be a legitimate academic discipline (like philosophy) that applies systematic techniques to the acquisition of information. For example, philosophers may differ on what they consider to be ethical behavior and may support their positions through logical argument. However, they lack any empirical test through which one view or another might be supported, and so the question of what is ethical cannot be addressed through scientific means.

Pseudoscience is another animal altogether. The term **pseudoscience** literally means "false science." According to Robert Carroll (2006), "pseudoscience is [a] set of ideas based on theories put forth as scientific when they are not scientific (<http://skepdic.com/pseudosc.html>)." It is important to note that true science and pseudoscience differ more in degree than in kind, with blurred boundaries between them (Lilienfeld, Lynn, & Lohr, 2003). What this means is that science and pseudoscience

share many characteristics. For example, both may attempt to provide support for an idea. However, the methods of pseudoscience do not have the same rigor or standards required of a true science. Some notorious examples of pseudoscience include phrenology (determining personality by reading the bumps on one's head), eye movement desensitization and reprocessing therapy (EMDR—moving one's eyes back and forth rapidly while thinking about a problem), and astrology (using the position of the stars and planets to explain behavior and predict the future).

Scott Lilienfeld (2005) lists several qualities that define a pseudoscience:

- ♦ Using situation-specific hypotheses to explain away falsification of a pseudoscientific idea or claim;
- ♦ Having no mechanisms for self-correction and consequent stagnation of ideas or claims;
- ♦ Relying on confirming one's beliefs rather than disconfirming them;
- ♦ Shifting the burden of proof to skeptics and critics away from the proponent of an idea or a claim;
- ♦ Relying on nonscientific anecdotal evidence and testimonials to support an idea or claim;
- ♦ Avoiding the peer review process that would scientifically evaluate ideas and claims;
- ♦ Failing to build on an existing base of scientific knowledge;
- ♦ Using impressive-sounding jargon that lends false credibility to ideas and claims;
- ♦ Failing to specify conditions under which ideas or claims would not hold true.

Lilienfeld points out that not one criterion from the above list is sufficient to classify an idea or claim as pseudoscientific. However, the greater the number of the aforementioned qualities an idea or claim possesses, the more confident you can be that the idea or claim is based on pseudoscience and not legitimate science.

Rory Coker (2007) provides a nice contrast between a true science and a pseudoscience. He identifies several crucial differences between science and pseudoscience that can help you assess whether an idea or claim is truly scientific or based on pseudoscientific beliefs. This contrast is shown in Table 1-1. Coker also suggests several additional characteristics of pseudoscience. First, pseudoscience often is unconcerned with facts and “spouts” dubious facts when necessary. Second, what research is conducted on an idea or claim is usually sloppy and does not include independent investigations to check its sources. Third, pseudoscience inevitably defaults to absurd explanations when pressed for an explanation of an idea or claim. Fourth, by leaving out critical facts pseudoscience creates mysteries that are difficult to solve. The full list of these and other characteristics of pseudoscience can be found at <http://www.quackwatch.org/01QuackeryRelatedTopics/pseudo.html>.

**TABLE 1-1 Distinguishing Science From Pseudoscience**

SCIENCE	PSEUDOSCIENCE
Findings published in peer-reviewed publications using standards for honesty and accuracy aimed at scientists.	Findings disseminated to general public via sources that are not peer reviewed. No prepublication review for precision or accuracy.
Experiments must be precisely described and be reproducible. Reliable results are demanded.	Studies, if any, are vaguely defined and cannot be reproduced easily. Results cannot be reproduced
Scientific failures are carefully scrutinized and studied for reasons for failure.	Failures are ignored, minimized, explained away, rationalized, or hidden.
Over time and continued research, more and more is learned about scientific phenomena.	No underlying mechanisms are identified and no new research is done. No progress is made and nothing concrete is learned.
Idiosyncratic findings and blunders “average out” and do not affect the actual phenomenon under study.	Idiosyncratic findings and blunders provide the only identifiable phenomena.
Scientists convince others based on evidence and research findings, making the best case permitted by existing data. Old ideas discarded in the light of new evidence.	Attempts to convince based on belief and faith rather than facts. Belief encouraged in spite of facts, not because of them. Ideas never discarded, regardless of the evidence.
Scientist has no personal stake in a specific outcome of a study.	Serious conflicts of interest. Pseudoscientist makes his or her living off of pseudoscientific products or services.

Based on information obtained from Coker (2007). <https://webspace.utexas.edu/cokerwr/www/index.html/distinguish.htm>

## Scientific Explanations

Contrast pseudoscience with how a true science operates. A true science attempts to develop scientific explanations to explain phenomena within its domain. Simply put, a **scientific explanation** is an explanation based on the application of accepted scientific methods. Scientific explanations differ in several important ways from non-scientific and pseudoscientific explanations that rely more on common sense or faith. Let's take a look at how science approaches a question like the effectiveness of EMDR therapy.

EMDR therapy was developed by Francine Shapiro. Shapiro noticed that when she was experiencing a disturbing thought her eyes were involuntarily moving rapidly. She noticed further that when she brought her eye movements under voluntary

control while thinking a traumatic thought, anxiety was reduced (Shapiro, 1989). Based on her experience, Shapiro proposed EMDR as a new therapy for individuals suffering from posttraumatic stress disorder (PTSD). Shapiro speculated that traumatic events “upset the excitatory/inhibitory balance in the brain, causing a pathological change in the neural elements” (Shapiro, 1989, p. 216). Shapiro speculated that the eye movements used in EMDR coupled with traumatic thoughts restored the neural balance and reversed the brain pathology caused by the trauma. In short, eye movements were believed to be central to the power of EMDR to bring about rapid and dramatic reductions in PTSD symptoms.

Shapiro (1989) provided some evidence for the effectiveness of EMDR therapy in the form of a case study. Based on her research and her case studies, Shapiro concluded that EMDR was a unique, effective new therapy for PTSD. Other researchers did not agree. They pointed out that Shapiro’s (and evidence provided by others) was based on flawed research. Because EMDR was rapidly gaining popularity, scientists began to test rigorously the claims made by advocates of EMDR. Two researchers, George Renfrey and C. Richard Spates (1994), set out to test systematically whether eye movements were, in fact, a necessary component of EMDR therapy. Their study provides an excellent example of how scientists go about their business of uncovering true scientific explanations.

In their experiment Renfrey and Spates “deconstructed” the EMDR technique into its components. Patients with PTSD were randomly assigned to one of three conditions in the study. Some patients were assigned to a standard EMDR condition. Other patients were assigned to an automated EMDR condition in which eye movements were induced by having patients shift their eyes back and forth between two alternating lights. The final group of patients was assigned to a no eye movement group in which the patients fixated their eyes on a stationary light. In all three conditions all of the other essential elements of EMDR therapy (visualizing and thinking about a traumatic event) were maintained. Measures of heart rate and anxiety were obtained from patients. Renfrey and Spates found that there was no difference between the three treatment groups on any of the measures, leading them to conclude that “eye movements are not an essential component of the intervention” (Renfrey & Spates, 1994, p. 238). Subsequent research confirmed this conclusion (Davidson & Parker, 2001).

In contrast to nonsense and pseudoscience, a true science attempts to develop scientific explanations for behavior through the application of the scientific method and specific scientific research designs, just as Renfrey and Spates (1994) did when they tested the role of eye movements in EMDR therapy. What sets a true scientific explanation apart from nonscientific and pseudoscientific explanations is that a scientific explanation is a tentative explanation, based on objective observation and logic, that can be empirically tested. Scientific explanations are the only ones accepted by scientists because they have a unique blend of characteristics that sets them apart from other explanations. Let’s take a look at those characteristics next.

**Scientific Explanations Are Empirical** An explanation is *empirical* if it is based on the evidence of the senses. To qualify as scientific, an explanation must be based

on objective and systematic observation, often carried out under carefully controlled conditions. The observable events and conditions referred to in the explanation must be capable of verification by others.

**Scientific Explanations Are Rational** An explanation is *rational* if it follows the rules of logic and is consistent with known facts. If the explanation makes assumptions that are known to be false, commits logical errors in drawing conclusions from its assumptions, or is inconsistent with established fact, then it does not qualify as scientific.

**Scientific Explanations Are Testable** A scientific explanation should either be verifiable through direct observation or lead to specific predictions about what should occur under conditions not yet observed. An explanation is *testable* if confidence in the explanation could be undermined by a failure to observe the predicted outcome. One should be able to imagine outcomes that would disprove the explanation.

**Scientific Explanations Are Parsimonious** Often more than one explanation is offered for an observed behavior. When this occurs, scientists prefer the **parsimonious explanation**, the one that explains behavior with the fewest number of assumptions.

**Scientific Explanations Are General** Scientists prefer explanations of broad explanatory power over those that “work” only within a limited set of circumstances.

**Scientific Explanations Are Tentative** Scientists may have confidence in their explanations, but they are nevertheless willing to entertain the possibility that an explanation is faulty. This attitude was strengthened in the past century by the realization that even Newton’s conception of the universe, one of the most strongly supported views in scientific history, had to be replaced when new evidence showed that some of its predictions were wrong.

**Scientific Explanations Are Rigorously Evaluated** This characteristic derives from the other characteristics listed, but it is important enough to deserve its own place in our list. Scientific explanations are constantly evaluated for consistency with the evidence and with known principles, for parsimony, and for generality. Attempts are made to extend the scope of the explanation to cover broader areas and to include more factors. As plausible alternatives appear, these are pitted against the old explanations in a continual battle for the “survival of the fittest.” In this way, even accepted explanations may be overthrown in favor of views that are more general, more parsimonious, or more consistent with observation.

## QUESTIONS TO PONDER

---

1. How do science, nonscience, and pseudoscience differ?
2. What are the defining characteristics of pseudoscience?
3. What are the main characteristics of scientific explanations? (Describe each.)

## Commonsense Explanations Versus Scientific Explanations

During the course of everyday experience, we develop explanations of the events we see going on around us. Largely, these explanations are based on the limited information available from the observed event and what our previous experience has told us is true. These rather loose explanations can be classified as **commonsense explanations** because they are based on our own sense of what is true about the world around us. Of course, scientific explanations and commonsense explanations have something in common: They both start with an observation of events in the real world. However, the two types of explanations differ in the level of proof required to support the explanation. Commonsense explanations tend to be accepted at face value, whereas scientific explanations are subjected to rigorous research scrutiny.

Take the case of Jerrod Miller, a Black man who was shot by a White off-duty police officer named Darren Cogoni in February 2005. Many in the Black community believed that Cogoni's behavior was racially motivated. The implication was that if Miller had been White, Cogoni would not have shot at him. That a police officer's racial prejudice might make him or her more quick to pull trigger on a minority suspect might seem to be a viable explanation for what happened in the Jerrod Miller case. Although this explanation may have some intuitive appeal, several factors disqualify it as a scientific explanation at this point.

First, the "racism" explanation was not based on careful, systematic observation. Instead, it was based on what people *believe* to be true of the relationship between race and a police officer's behavior. Consequently, the explanation may have been derived from biased, incomplete, or limited evidence (if from any evidence at all). Second, it was not examined to determine its consistency with other available observations. Third, no effort was made to evaluate it against plausible alternative explanations. Fourth, no predictions were derived from the explanation and tested. Fifth, no attempt was made to determine how well the explanation accounted for similar behavior in a variety of other circumstances. The explanation was accepted simply because it appeared to make sense of Cogoni's behavior and was consistent with pre-existing beliefs about how the police treat Black suspects.

Because commonsense explanations are not rigorously evaluated, they are likely to be incomplete, inconsistent with other evidence, lacking in generality, and probably wrong. This is certainly the case with the "racism" explanation. Most individuals who harbor racial prejudices do not behave aggressively toward minority-group members. Other factors must also contribute.

Although commonsense explanations may "feel right" and give us a sense that we understand a behavior, they may lack the power to apply across a variety of apparently similar situations. To see how commonsense explanations may fail to provide a truly general account of behavior, consider the following event.

Late in December 1903, a fire started in the crowded Iroquois Theater of Chicago, and 602 people lost their lives. Of interest to psychologists is not the fact that 602 people died, *per se*, but rather the circumstances that led to many of the deaths. Many of the victims were not directly killed by the fire. Rather, they were trampled to death in the panic that ensued in the first few minutes after the fire started. In his classic book *Social Psychology*, Brown (1965) reproduced an account

of the event provided by Eddie Foy, a famous comedian of the time. According to Foy's account,

[I]t was inside the house that the greatest loss of life occurred, especially on the stairways leading down from the second balcony. Here most of the dead were trampled or smothered. . . . In places on the stairways, particularly where a turn caused a jam, bodies were piled seven or eight deep. (Brown, 1965, p. 715)

As a student of psychology, you may already be formulating explanations for why normally rational human beings would behave mindlessly in this situation. Clearly, many lives would have been saved had the patrons of the Iroquois Theater filed out in an orderly fashion. How would you explain the tragedy?

A logical and "obvious" answer is that the patrons believed their lives to be in danger and wanted to leave the theater as quickly as possible. In this view, the panic inside the theater was motivated by a desire to survive.

Notice that the explanation at this point is probably adequate to explain the crowd behavior under the specific conditions inside the theater and perhaps to explain the same behavior under other life-threatening conditions. However, the explanation is probably too situation specific to serve as a general scientific explanation of irrational crowd behavior. It cannot explain, for example, the following incident.

On December 10, 1979, a crowd of young people lined up outside a Cincinnati arena to wait for the doors to open for a concert by the rock group the Who. As the doors opened, the crowd surged ahead. Eleven people were trampled to death even though the conditions were certainly not life-threatening. In fact, the identifiable reward in this situation was obtaining a good seat at an open-seating concert.

Clearly, the explanation for irrational crowd behavior at the Chicago theater cannot be applied to the Cincinnati tragedy. People were not going to die if they failed to get desirable seats at the concert. What seemed a reasonable explanation for irrational crowd behavior in the Iroquois Theater case must be discarded.

You must look for common elements to explain such similar yet diverse events. In both situations, the available rewards were perceived to be limited. A powerful reward (avoiding pain and death) in the Iroquois Theater undoubtedly was perceived as attainable only for a brief time. Similarly, in Cincinnati the perceived reward (a seat close to the stage), although not essential for survival, was also available for a limited time only. In both cases, apparently irrational behavior resulted as large numbers of people individually attempted to maximize the probability of obtaining the reward.

The new tentative explanation for the irrational behavior now centers on the perceived availability of rewards rather than situation-specific variables. This new tentative explanation has been tested in research and has received some support.

As these examples illustrate, simple commonsense explanations may not apply beyond the specific situations that spawned them. The scientist interested in irrational crowd behavior would look for a more general concept (such as perceived availability of rewards) to explain observed behavior. That is not to say that simple, obvious explanations are always incorrect. However, when you are looking for an explanation that transcends situation-specific variables, you often must look beyond simple, commonsense explanations.

## Belief-Based Explanations Versus Scientific Explanations

Explanations for behavior often arise not from common sense or scientific observation but from individuals or groups who (through indoctrination, upbringing, or personal need) have accepted on faith the truth of their beliefs. You may agree or disagree with those beliefs, but you should be aware that explanations offered by science and **belief-based explanations** are fundamentally different.

Explanations based on belief are accepted because they come from a trusted source or appear to be consistent with the larger framework of belief. No evidence is required. If evidence suggests that the explanation is incorrect, then the evidence is discarded or reinterpreted to make it appear consistent with the belief. For example, certain religions hold that Earth was created only a few thousand years ago. The discovery of fossilized remains of dinosaurs and other creatures (apparently millions of years old) challenged this belief. To explain the existence of these remains, people defending the belief suggest that fossils are actually natural rock formations that resemble bones or that the fossils are the remains of the victims of the Great Flood. Thus, rather than calling the belief into question, apparently contrary evidence is interpreted to appear consistent with the belief.

This willingness to apply a different *post hoc* (after-the-fact) explanation to reconcile the observations with belief leads to an unparsimonious patchwork quilt of explanations that lacks generality, fails to produce testable predictions about future findings, and often requires that one assumes the common occurrence of highly unlikely events. Scientific explanations of the same phenomena, in contrast, logically organize the observed facts by means of a few parsimonious assumptions and lead to testable predictions.

Nowhere is the contrast between these two approaches more striking than in the current debate between evolutionary biologists and the so-called creation scientists, whose explanation for fossils was previously mentioned. To take one example, consider the recent discoveries based on gene sequencing, which reveal the degree of genetic similarity among various species. These observations and some simple assumptions about the rate of mutation in the genetic material allowed biologists to develop “family trees” indicating how long ago the various species separated from one another. The trees drawn up from the gene-sequencing data agree amazingly well with and to a large degree were predicted by the trees assembled from the fossil record. In contrast, because creationists assume that all animals alive today have always had their current form and that fossils represent the remains of animals killed in the Great Flood, their view could not have predicted relationships found in the genetic material. Instead, they must invent yet another *post hoc* explanation to make these new findings appear consistent with their beliefs.

In addition to the differences described thus far, scientific and belief-based explanations also differ in tentativeness. Whereas explanations based on belief are simply *assumed* to be true, scientific explanations are accepted because they are consistent with existing objective evidence and have survived rigorous testing against plausible alternatives. Scientists accept the possibility that better explanations may turn up or that new tests may show that the current explanation is inadequate.



Scientific explanations also differ from belief-based explanations in the subject areas for which explanations are offered. Whereas explanations based on belief may seek to answer virtually any question, scientific explanations are limited to addressing those questions that can be answered by means of objective observations. For example, what happens to a person after death and why suffering exists in the world are explained by religion, but such questions remain outside the realm of scientific explanation. No objective tests or observations can be performed to answer these questions within the confines of the scientific method. Science offers no explanation for such questions, and you must rely on faith or belief for answers. However, for questions that can be settled on the basis of objective observation, scientific explanations generally have provided more satisfactory and useful accounts of behavior than those provided by a priori belief.

### QUESTIONS TO PONDER

---

1. How do scientific and commonsense explanations differ?
2. How do belief-based and scientific explanations differ?
3. What kinds of questions do scientists refrain from investigating? Why do scientists refrain from studying these issues?

### WHEN SCIENTIFIC EXPLANATIONS FAIL

---

Scientific explanation is preferable to other kinds of explanation when scientific methods can be applied. Using a scientific approach maximizes the chances of discovering the best explanation for an observed behavioral phenomenon. Despite the application of the most rigorous scientific methods, instances do occur in which the explanation offered by a scientist is not valid. Scientific explanations are sometimes flawed. Understanding some of the pitfalls inherent to developing scientific explanations will help you avoid arriving at flawed or incorrect explanations for behavior.

#### Failures Due to Faulty Inference

Explanations may fail because developing them involves an inference process. We make observations and then infer the causes for the observed behavior. This inference process always involves the danger of incorrectly inferring the underlying mechanisms that control behavior.

The problem of faulty inference is illustrated in a satirical book by David Macaulay (1979) called *Motel of the Mysteries*. In this book, a scientist (Howard Carson) uncovers the remnants of our civilization 5,000 years from now. Carson unearths a motel and begins the task of explaining what our civilization was like, based on the artifacts found in the motel.

Among the items unearthed were various bathroom plumbing devices: a plunger, a showerhead, and a spout. These items were assumed by Carson to be musical instruments. The archaeologist describes the items as follows:

The two trumpets [the showerhead and spout] . . . were found attached to the wall of the inner chamber at the end of the sarcophagus. They were both coated with a silver substance similar to that used on the ornamental pieces of the metal animals. Music was played by forcing water from the sacred spring through the trumpets under great pressure. Pitch was controlled by a large silver handle marked HC. . . . The [other] instrument [the plunger] is probably of the percussion family, but as yet the method of playing it remains a mystery. It is, however, beautifully crafted of wood and rubber. (Macaulay, 1979, p. 68)

By hypothesizing that various plumbing devices served as ceremonial musical instruments, Macaulay's archaeologist has reached a number of inaccurate conclusions. Although the *Motel of the Mysteries* example is pure fiction, real-life examples of inference gone wrong abound in science, and psychology is no exception. R. E. Fancher (1985) described the following example in his book *The Intelligence Men: Makers of the IQ Controversy*. During World War I, the U.S. Army administered group intelligence tests under the direction of Robert Yerkes. More than 1.75 million men had taken either the Alpha or Beta version of the test by the end of the war and provided an excellent statistical sample from which conclusions could be drawn about the abilities of U.S. men of that era.

The results were shocking. Analysis of the data revealed that the average army recruit had a mental age of 13 years—3 years below the “average adult” mental age of 16 and only 1 year above the upper limit for morosity. Fancher described Yerkes's interpretation as follows:

Rather than interpreting his results to mean that there was something wrong with the standard, or that the army scores had been artificially depressed by . . . the failure to re-test most low Alpha scorers on Beta, as was supposed to have been the case, Yerkes asserted that the “native intelligence” of the average recruit was shockingly low. The tests, he said, were “originally intended, and now definitely known, to measure native intellectual ability. They are to some extent influenced by educational acquirement, but in the main the soldier's inborn intelligence and not the accidents of environment determined his mental rating or grade.” Accordingly, a very substantial proportion of the soldiers in the U.S. Army were actually morons. (1985, p. 127)

In fact, Yerkes's assertions about the tests were not in any sense established, and indeed the data provided evidence against Yerkes's conclusion. For example, poorly educated recruits from rural areas scored lower than their better-educated city cousins. Yerkes's tests had failed to consider the differences in educational opportunities among recruits. As a result, Yerkes and his followers inappropriately concluded that the average intellectual ability of Americans was deteriorating.

In the Yerkes example, faulty conclusions were drawn because the conclusions were based on unfounded assumptions concerning the ability of the tests to

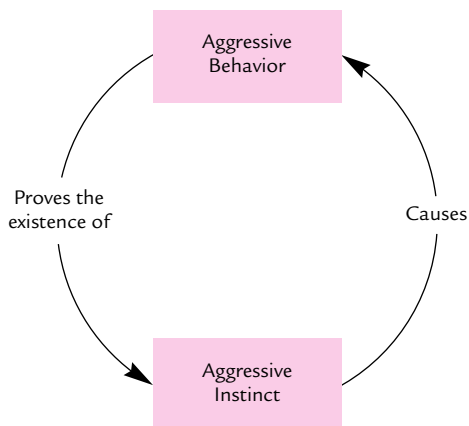
unambiguously measure intelligence. The researchers failed to consider possible *alternative explanations* for observed effects. Although the intelligence of U.S. Army recruits may in fact have been distressingly low, an alternative explanation centering on environmental factors such as educational level would have been equally plausible. These two rival explanations (real decline in intelligence versus lack of educational experience) should have been subjected to the proper tests to determine which was more plausible. Later, this book discusses how developing, testing, and eliminating such rival hypotheses are crucial elements of the scientific method.

### Pseudoexplanations

Failing to consider alternative explanations is not the only danger waiting to befall the unwary scientist. In formulating valid scientific explanations for behavioral events, it is important to avoid the trap of **pseudoexplanation**. In seeking to provide explanations for behavior, psychologists sometimes offer positions, theories, and explanations that do nothing more than provide an alternative label for the behavioral event. One notorious example was the attempt to explain aggression with the concept of an instinct. According to this position, people (and animals) behave aggressively because of an aggressive instinct. Although this explanation may have intuitive appeal, it does not serve as a valid scientific explanation.

Figure 1-1 illustrates the problem with such an explanation. Notice that the observed behavior (aggression) is used to prove the existence of the aggressive instinct. The concept of instinct is then used to explain the aggressive behavior.

This form of reasoning is called a **circular explanation**, or **tautology**. It does not provide a true explanation but rather merely provides another label (instinct) for a class of observed behavior (aggression). Animals are aggressive because they have aggressive instincts. How do we know they have aggressive instincts? Because they are aggressive! Thus, all we are saying is that animals are aggressive because of a tendency to behave aggressively. Obviously, this is not an explanation.



**FIGURE 1-1** A circular explanation. The observed behavior is “explained” by a concept, but the behavior itself is used as proof of the existence of the explanatory concept.

You might expect only novice behavioral scientists to be prone to using pseudoexplanations. However, even professional behavioral scientists have proposed “explanations” for behavioral phenomena that are really pseudoexplanations. In a 1970 article, Martin Seligman proposed a continuum of preparedness to help explain why an animal can learn some associations easily (such as between taste and illness) and other associations only with great difficulty (such as between taste and electric shock).

According to Seligman’s analysis, the animal may be biologically prepared to learn some associations (those learned quickly) and contraprepared to learn others (those learned slowly, if at all). Thus, some animals may have difficulty acquiring an association between taste and shock because they are contraprepared by evolution to associate the two.

As with the use of instinct to explain aggression, the continuum-of-preparedness notion seems intuitively correct. Indeed, it does serve as a potentially valid explanation for the observed differences in learning rates. But it does not qualify as a true explanation as it is stated. Refer to Figure 1-1 and substitute “quickly or slowly acquired association” for “aggressive behavior” and “continuum of preparedness” for “aggressive instinct.” As presently stated, the continuum-of-preparedness explanation is circular: Animals learn a particular association with difficulty because they are contraprepared to learn it. How do you know they are contraprepared? You know because they have difficulty learning.

How can you avoid falling into the trap of proposing and accepting pseudoexplanations? When evaluating a proposed explanation, ask yourself whether or not the researcher has provided *independent measures* of the behavior of interest (such as difficulty learning an association) *and* the proposed explanatory concept (such as the continuum of preparedness). For example, if you could find an independent measure of preparedness that does *not* involve the animal’s ability to form an association, then the explanation in terms of preparedness would qualify as a true explanation. If you can determine the animal’s preparedness only by observing its ability to form a particular association, the proposed explanation is circular. Rather than explaining the differing rates of learning, the statement actually serves only to define the types of preparedness.

Developing independent measures for the explanatory concept and the behavior to be explained may not be easy. For example, in the continuum-of-preparedness case, it may take some creative thought to develop a measure of preparedness that is independent of the observed behavior. The same is true for the concept of an instinct.

As these examples have illustrated, even scientific explanations may fail. However, you should not conclude that such explanations are no better than those derived from other sources. Living, behaving organisms are complex systems whose observable workings provide only clues to their inner processes. Given the available evidence, you make your best guess. It should not be surprising that these guesses are often wrong. As these conjectures are evaluated against new evidence, even the failures serve to rule out plausible alternatives and to prepare the way for better guesses. As a result, science has a strong tendency to converge on valid explanations as research progresses. Such progress in understanding is a hallmark of the scientific method.

## QUESTIONS TO PONDER

---

1. How can faulty inference invalidate a scientific explanation?
2. What are pseudoexplanations, and how do you avoid them?

## METHODS OF INQUIRY

---

Before a scientist can offer valid and general explanations for behavior, he or she must gather information about the behavior of interest. Knowledge about behavior can be acquired by several methods, including the method of authority, the rational method, and the scientific method.

### The Method of Authority

After reading about the Iroquois Theater tragedy, you might make a trip to your local public or university library or call your former social psychology professor in search of information to help explain the irrational behavior inside the theater. When you use expert sources (whether books or people), you are using the **method of authority**. Using the method of authority involves consulting some source that you consider authoritative on the issue in question (e.g., consulting books, television, religious leaders, scientists).

Although useful in the early stages of acquiring knowledge, the method of authority does not always provide valid answers to questions about behavior for at least two reasons. First, the source that you consult may not be truly authoritative. Some people (such as Lucy in the *Peanuts* comic strip) are more than willing to give you their “expert” opinions on any topic, no matter how little they actually know about it (writers are no exception). Second, sources often are biased by a particular point of view. A sociologist may offer a different explanation for the Iroquois Theater tragedy from the one offered by a behaviorally oriented psychologist. For these reasons, the method of authority by itself is not adequate for producing reliable explanations.

Although the method of authority is not the final word in the search for explanations of behavior, the method does play an important role in the acquisition of scientific knowledge. Information that you obtain from authorities on a topic can familiarize you with the problem, the available evidence, and the proposed explanations. With this information, you could generate new ideas about causes of behavior. However, these ideas must then be subjected to rigorous scientific scrutiny rather than being accepted at face value.

### The Rational Method

René Descartes proposed in the 17th century that valid conclusions about the universe could be drawn through the use of pure reason, a doctrine called *rationalism*. This proposal was quite revolutionary at the time because most scholars of the day relied heavily on the method of authority to answer questions. Descartes’ method began with skepticism, a willingness to doubt the truth of every belief. Descartes

noted, as an example, that it was even possible to doubt the existence of the universe. What you perceive, he reasoned, could be an illusion. Could you prove otherwise?

After establishing doubt, Descartes moved to the next stage of his method: the search for “self-evident truths,” statements that *must* be true because to assume otherwise would contradict logic. Descartes reasoned that if the universe around him did not really exist, then perhaps he himself also did not exist. It was immediately obvious to Descartes that this idea contradicted logic—it was self-evidently true that if he did not exist, he certainly could not be thinking about the question of his own existence. And it was just as self-evidently true that he was indeed thinking.

These two self-evident truths can be used as assumptions from which deductive logic will yield a firm conclusion:

Assumption 1: Something that thinks must exist.

Assumption 2: I am thinking.

Conclusion: I exist.

Using only his powers of reasoning, Descartes had identified two statements whose truth logically cannot be doubted, and from them he was able to deduce a conclusion that is equally bulletproof. It is bulletproof because, if the assumptions are true and you make no logical errors, deduction guarantees the truth of the conclusion. By the way, this particular example of the use of his method was immortalized by Descartes in the declaration “*Cogito, ergo sum*” (Latin for “I think, therefore I am”). If you’ve heard that phrase before and wondered what it meant, now you know.

Descartes’ method came to be called the **rational method** because it depends on logical reasoning rather than on authority or the evidence of one’s senses. Although the method satisfied Descartes, we must approach “knowledge” acquired in this way with caution. The power of the rational method lies in logically deduced conclusions from self-evident truths. Unfortunately, precious few self-evident truths can serve as assumptions in a logical system. If one (or both) of the assumptions used in the deduction process is incorrect, the logically deduced conclusion will be invalid.

Because of its shortcomings, the rational method is not used to develop scientific explanations. However, it still plays an important role in science. The tentative ideas that we form about the relationship between variables are often deduced from earlier assumptions. For example, having learned that fleeing from a fire or trying to get into a crowded arena causes irrational behavior, we may deduce that “perceived availability of reinforcers” (escaping death or getting a front-row seat) is responsible for such behavior. Rather than accepting such a deduction as correct, however, the scientist puts the deduction to empirical test.

## The Scientific Method

Braithwaite (1953) proposed that the function of a science is to “establish general laws covering the behavior of the empirical events with which the science in question is concerned” (p. 1). According to Braithwaite, a science should allow us to fuse together information concerning separately occurring events and to make reliable predictions about future, unknown events. One goal of psychology is to establish

general laws of behavior that help explain and predict behavioral events that occur in a variety of situations.

Although explanations for behavior and general laws cannot be adequately formulated by relying solely on authoritative sources and using deductive reasoning, these methods (when combined with other features) form the basis for the most powerful approach to knowledge yet developed: the **scientific method**. This method comprises a series of four cyclical steps that you can repeatedly execute as you pursue the solution to a scientific problem (Yaremko et al., 1982, p. 212). These steps are (1) observing a phenomenon, (2) formulating tentative explanations or statements of cause and effect, (3) further observing or experimenting (or both) to rule out alternative explanations, and (4) refining and retesting the explanations.

**Observing a Phenomenon** The starting point for using the scientific method is to observe the behavior of interest. This first step is essentially what Cialdini (1994) called “scouting” in which some behavior or event catches your attention. These preliminary observations of behavior and of potential causes for that behavior can take a variety of forms. In the case of the effects of distraction on driving ability, your initial musings about Bailey Goodman’s accident may have led you to think more carefully about the role of distraction on the ability to perform a complex task. Your curiosity might have been further piqued when divided attention was discussed in your cognitive psychology class or when you read about another case where cell phone distraction was a suspected cause of an accident. Or you might even have known someone who was nearly killed in an accident while talking on his cell phone and driving his car at the same time. In any of these cases, your curiosity might be energized so that you begin to formulate hypotheses about what factors affect the behavior you have observed.

Through the process of observation, you identify variables that appear to have an important influence on behavior. A **variable** is any characteristic or quantity that can take on two or more values. For example, whether a participant is talking on a cell phone or not while doing a simulated driving task is a variable. Remember that in order for something to be considered a variable it must be capable of taking on *at least* two values (e.g., talking on a cell phone or not talking on a cell phone). A characteristic or quantity that takes on only one value is known as a *constant*.

**Formulating Tentative Explanations** After identifying an interesting phenomenon to study, your next step is to develop one or more tentative explanations that seem consistent with your observations. In science these tentative explanations often include a statement of the relationship between two or more variables. That is, you tentatively state the nature of the relationship between variables that you expect to uncover with your research. The tentative statement that you offer concerning the relationship between your variables of interest is called a **hypothesis**. It is important that any hypothesis you develop be testable with empirical research.

As an example of formulating a hypothesis, consider the issue of the relationship between talking on a cell phone and driving. After your preliminary observations, you might formulate the following hypothesis: A person is more likely to make driving errors when talking on a cell phone than not talking on a cell phone.

Notice that the hypothesis links two variables (talking or not talking on a cell phone and errors on a driving task) by a statement indicating the expected relationship between them. In this case, the relationship expected is that talking on a cell phone will increase errors on the simulated driving task. Research hypotheses often take the form of a statement of how changes in the value of one variable (rate of suspect) will affect the value of the other variable (a decision to shoot).

**Further Observing and Experimenting** When Cialdini (1994) talked about “trapping” effects, he was referring to the process of designing empirical research studies to isolate the relationship between the variables chosen for study. Up to the point of developing a hypothesis, the scientific method does not differ markedly from other methods of acquiring knowledge. At this point, all you have done is to identify a problem to study and develop a hypothesis based on some initial observation. The scientific method, however, does not stop here. The third step in the scientific method marks the point at which the scientific method differs from other methods of inquiry. Unlike the other methods of inquiry, the scientific method demands that further observations be carried out to test the validity of any hypotheses that you develop. In other words, “a-trapping we shall go.”

What exactly is meant by “making further observations”? The answer to this question is what the scientific method is all about. After formulating your hypothesis, you design a research study to test the relationship that you proposed. This study can take a variety of forms. It could be a *correlational study* in which you simply measure two or more variables and look for a relationship between them (see Chapter 4), a *quasi-experimental study* in which you take advantage of some naturally occurring event or preexisting conditions, or an *experiment* in which you systematically manipulate a variable and look for changes in the value of another that occur as a result (see Chapters 10–12).

In this case, you decide to design an experiment in which you systematically manipulate whether a person talks on a cell phone and observe the number of errors made on a simulated driving task.

**Refining and Retesting Explanations** The final step in the scientific method is the process of refinement and retesting. As an example of this process, imagine that you found that individuals are more likely to make driving errors when talking on a cell phone. Having obtained this result, you would probably want to explore the phenomenon further: Would talking on a cell phone cause more driving errors than having a conversation with a passenger? A refined research hypothesis might take the following form: Individuals are more likely to make driving errors when talking on a cell phone than when having a conversation with a passenger.

This process of generating new, more specific hypotheses in the light of previous results illustrates the *refinement process*. Often, confirming a hypothesis with a research study leads to other hypotheses that expand on the relationships discovered, explore the limits of the phenomenon under study, or examine the causes for the relationship observed.



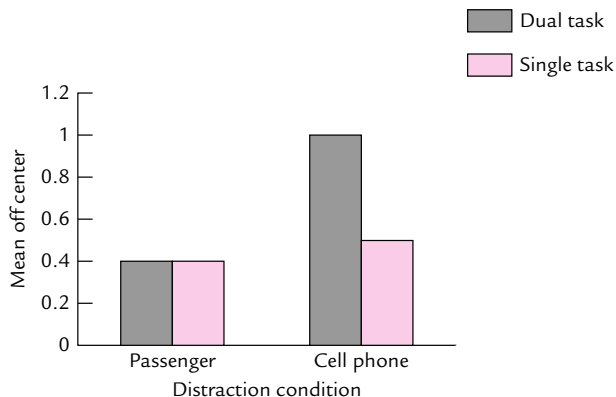
As you become more familiar with the process of conducting research, you will find that not all research studies produce affirmative results. That is, sometimes your research does not confirm your hypothesis. What do you do then? In some cases, you might completely discard your original hypothesis. In other cases, however, you might revise and retest your hypothesis. In the latter instance, you are using a strategy known as *retesting*. Keep in mind that any revised or refined hypothesis must be tested as rigorously as was the original hypothesis.

The scientific method requires a great deal of time making careful observations. Sometimes your observations don't confirm your hypothesis. Is the scientific method worth all the extra effort? In fact, the ability to discover that a relationship does *not* exist makes the scientific method the powerful tool that it is. By repeatedly checking and rechecking hypotheses in the ruthless arena of empirical testing, the scientist learns which ideas are worthy and which belong on the trash heap. No other method incorporates such a powerful check on the validity of its conclusions.

### The Scientific Method at Work: Talking on a Cell Phone and the Ability to Drive

Throughout this chapter, we've used the issue of the safety of talking on a cell phone while driving to illustrate how you might go about developing, testing, and refining a research hypothesis. As you may have suspected, the question has actually been the subject of scientific research, and we thought it might be helpful for you to see how an actual research study on this topic was carried out. The study that we chose for our example was conducted by Drews, Pasupathi, and Strayer (2008).

In their experiment, Drews et al. (2008) had college students perform a realistic simulated driving task. The simulated driving task required participants to navigate a 24-mile multilane road complete with on-off ramps, overpasses, and two-way traffic. The simulation required participants to merge into traffic, deal with other cars on the road, maneuver around slow-moving traffic, and regulate speed. Participants had to navigate the course and bring their "car" to a stop in a rest stop on the course. Two participants at a time took part in the experiment, one randomly assigned as the driver and the other either a passenger or someone talking to the driver on a cell phone. The "passenger" sat next to the driver and engaged the driver with a story in which the passenger had a close call. In the "cell phone" scenario the conversation was held via cell phone with the talker separated from the driver. Before completing the drive with either the cell phone or passenger conversation (dual task condition), participants performed the simulated driving task alone (single task condition) with no conversation taking place. Drews et al. collected four driving performance measures: (1) how well the drivers could stay in the center of the lane, (2) speed, (3) following distance, and (4) how successfully they completed the task (i.e., getting off the highway at the rest stop). The results for one of the measures are shown in Figure 1-2. As you can see, drivers who were having a conversation on a cell phone showed more deviation from the center of their lane than those having a conversation with a passenger. There was no difference in deviation from lane center when participants did the driving task alone.



**FIGURE 1-2** Results from the Drews, Pasupathi, & Strayer (2008) distracted driving experiment. Based on data provided by Drews, et al. (2008)

## QUESTIONS TO PONDER

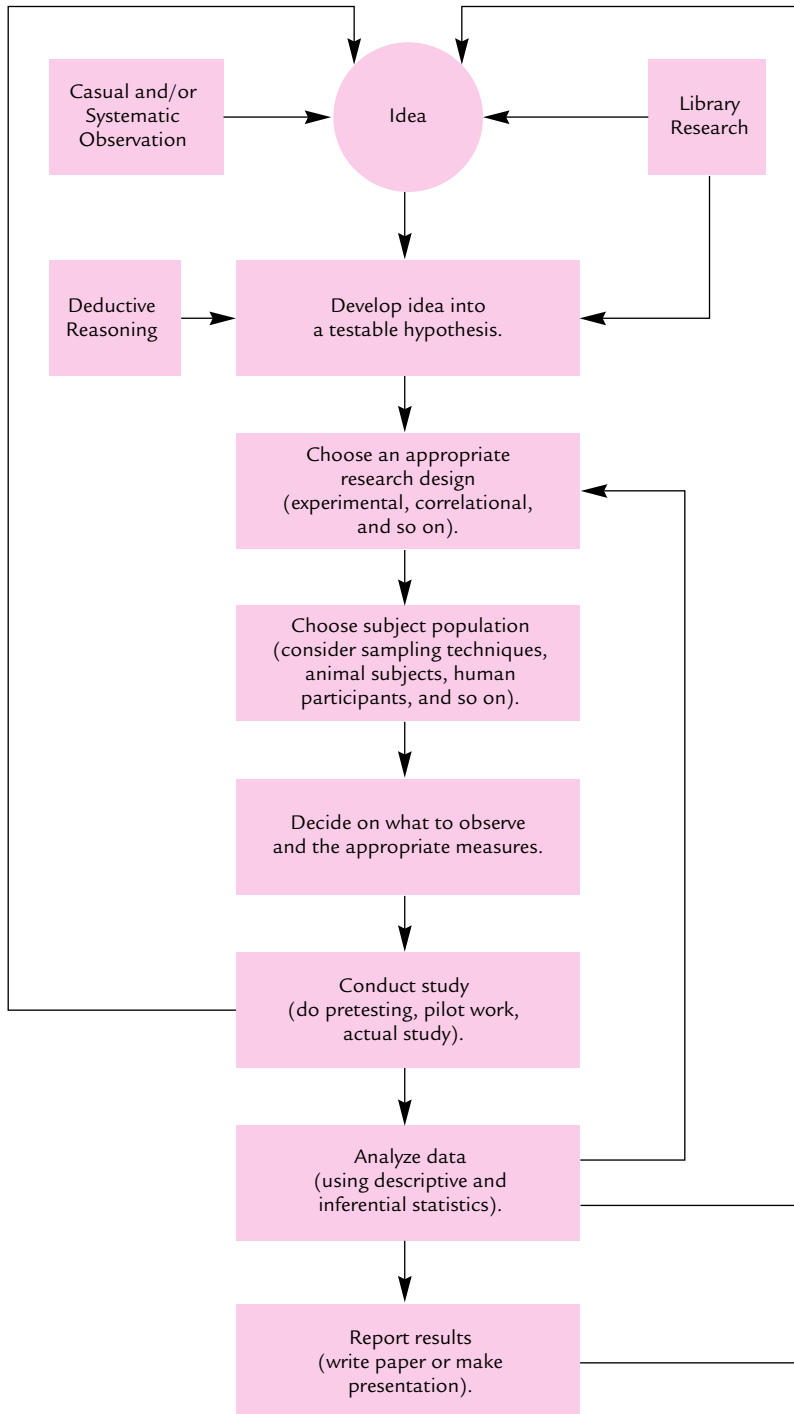
1. What are the defining characteristics and weaknesses of the method of authority and the rational method?
2. How are the method of authority and rational method used in science?
3. What are the steps involved in the scientific method?
4. Why is the scientific method preferred in science?

## The Steps of the Research Process

Scientists in the field of psychology adhere to the scientific method as the principal method for acquiring information about behavior. This is true whether the psychologist is a *clinical psychologist* evaluating the effectiveness of a new therapy technique or an *experimental psychologist* investigating the variables that affect memory. Of course, researchers in psychology adopt a wide variety of techniques in their quest for scientific knowledge.

From the inception of a research idea to the final report of results, the research process has several crucial steps. These steps are outlined in Figure 1-3. At each step you must make one or more important decision that will influence the direction of your research. Let's explore each of these steps and some of the decisions you must make.

**Developing a Research Idea and Hypothesis** The first step in the research process is to identify an issue that you want to study. There are many sources of research ideas (e.g., observing everyday behavior or reading scientific journals). Once you have identified a behavior to study, you must then state a research question in terms that will allow others to test it empirically. Many students of research have trouble at this point. Students seem to have little trouble identifying interesting, broadly defined behaviors to study (e.g., "I want to study memory"), but they have trouble isolating crucial variables that need to be explored.



**FIGURE 1-3** The research process. Arrows show the sequence of steps, along with feedback pathways.

To apply the scientific method, rationally, you must be able to state clearly the relationships that you expect to emerge in a research study. In other words, you must be able to formulate a precise, testable hypothesis. As noted in Figure 1-3, hypothesis development involves **deductive reasoning**, which involves deriving a specific hypothesis (in this case) from general ideas. For example, during your literature review you may have come across a theory about how memory operates. Using the general ideas developed in a theory, you may logically deduce that one variable (e.g., meaningfulness of the information to be learned) causes changes in a second (amount remembered). The specific statement connecting these two variables is your hypothesis.

**Choosing a Research Design** Once you have narrowed your research question and developed a testable hypothesis, you must next decide on a design or plan of attack for your research. As discussed in later chapters, a variety of options is available. For example, you must decide whether to do a correlational study (measure two or more variables and look for relationships among them) or an experimental study (manipulate a variable and look for concomitant changes in a second). Other important decisions at this point include where to conduct your study (in the laboratory or in the field) and how you are going to measure the behavior of interest.

With the preliminary decisions out of the way, you must consider a host of practical issues (equipment needs, preparation of materials, etc.). You might find it necessary to conduct a miniature version of your study, called a **pilot study**, to be sure your chosen procedures and materials work the way that you think they will.

**Choosing Subjects** Once you have designed your study and tested your procedures and materials, you need to decide whether to use human participants or animal subjects. You must decide how to obtain your subjects and how they will be handled in your study. You also must be concerned with treating your subjects in an ethical manner.

**Deciding on What to Observe and Appropriate Measures** Your next step is to decide exactly what it is you want to observe, which will be determined by the topic or issue that you have chosen to investigate. For example, if you were interested in the issue of the impact of media violence on children's aggression, you might interview parents who have noticed an increase in aggression after their children play violent video games. Or you might design an experiment similar to Drews et al.'s (2008) experiment to test the effects of distraction on driving ability. After choosing what to observe, you must next decide on the most appropriate way to measure the behavior of interest. For example, should you use the same measure that Drews and other experimenters used, or should you develop a new one?

**Conducting Your Study** Now you actually have your participants take part in your study. You observe and measure their behavior. Data are formally recorded for later analysis.

**Analyzing Your Results** After you have collected your data, you must summarize and analyze them. The analysis process involves a number of decisions. You can

analyze your data in several ways, and some types of data are better analyzed with one method than another. In most cases, you will probably calculate some descriptive statistics that provide a “nutshell” description of your data (such as averages and standard deviations) and inferential statistics that assess the reliability of your data (such as a *t* test).

**Reporting Your Results** After analyzing your data, you are nearing the final steps in the research process. You are now ready to prepare a report of your research. If your results were reliable and sufficiently important, you may want to publish them. Consequently, you would prepare a formal paper, usually in American Psychological Association (APA) style, and submit it to a journal for review. You also might decide to present your paper at a scientific meeting, in which case you prepare a brief abstract of your research for review.

**Starting the Whole Process Over Again** Your final report of your research is usually not the final step in your research. You may have achieved closure on (finished and analyzed) one research project. However, the results from your first study may raise more questions. These questions often serve as the seeds for a new study. In fact, you may want to replicate an interesting finding within the context of a new study. This possibility is represented in Figure 1-3 by the arrow connecting “Report results” with “Idea.”

## QUESTIONS TO PONDER

---

1. What are the steps involved in the research process?
2. What important decisions must be made at each step of the research process?

## SUMMARY

---

Although we are constantly trying to explain the behavior that we see around us, commonsense explanations of behavior often are too simplistic, situation specific, and frequently based on hearsay, conjecture, anecdote, or other unreliable sources. Scientific explanations are based on carefully made observations of behavior, rigorously tested against alternative explanations, and developed to provide the most general account that is applicable over a variety of situations. For these reasons, scientific explanations tend to be more valid and general than those provided by common sense.

The goal of the science of psychology is to build an organized body of knowledge about its subject matter and to develop explanations for phenomena within its domain. It is important to distinguish between a true science, nonscience, and pseudoscience because the quality of the information obtained depends on how it is acquired. The principal method used in a true science to build an organized body of knowledge and develop scientific explanations is research. Research involves three

steps: identifying a phenomenon to study, discovering information about that phenomenon, and developing explanations for the phenomenon. A useful analogy is to think of science as a hunting trip. First, you scout where you are going to hunt for prey (analogous to identifying a phenomenon to study). Second, you go hunting to trap your prey (analogous to discovering information and developing explanations).

Explanations for behavior also are provided by beliefs. Explanations provided by belief differ from scientific explanations in that they are considered absolutely true, whereas scientific explanations are always considered tentative. Consequently, when evidence conflicts with an explanation based on belief, the evidence is questioned. When evidence conflicts with a scientific explanation, the explanation is questioned. Although beliefs can provide answers to virtually any question, the scientific method can address only those questions that can be answered through observation.

Even explanations that sound scientific may fail because relationships are often inferred from observable events. The danger always exists that inferences are incorrect, despite being based on empirical data. An explanation also may fail if you do not use independent measures of the explanatory concept and the behavior to be explained. In such cases, you have a pseudoexplanation, which is only a new label for behavior.

There are many ways to acquire knowledge about behavior. With the method of authority, you acquire information from sources that you perceive to be expert on your topic of interest and use the information to develop an explanation for behavior. With the rational method, you deduce explanations from other sources of information. Although the method of authority and the rational method play important roles in the early stages of science, they are not acceptable methods for acquiring scientific knowledge. The scientific method is the only method accepted for the acquisition of scientific knowledge.

The four major steps of the scientific method are (1) observation of a phenomenon, (2) formation of tentative explanations or statements of cause and effect, (3) further observation or experimentation to rule out alternative explanations (or both), and (4) retesting and refinement of the explanations. The scientific method is also an attitude or a way of viewing the world. The scientist frames problems in terms of the scientific method.

The scientific method is translated into action by the research process. When performing research, you first choose a technique. Regardless of the technique chosen, research must follow the guidelines of the scientific method. The science of psychology is highly complex and diverse, and the goals of research vary from individual to individual. Some researchers, who are mainly interested in solving real-world problems, conduct applied research. Other scientists, mainly those interested in evaluating theoretical problems, conduct basic research. Even though basic and applied research are different to some extent, considerable overlap does exist. Some basic research problems have real-world applications, and some applied problems have some basic research undertones.

The research process involves a sequence of steps. At each step, important decisions affect the course of research and how you analyze and interpret data. The steps in the research process are (1) develop a research idea into a testable hypothesis, (2) choose a research design, (3) choose a subject or participant population,

(4) decide on what to observe and appropriate measures, (5) obtain subjects or participants for the study and conduct the study, (6) analyze results, and (7) report results. Often the results of research raise a host of new research ideas, which starts the whole research process over again.

## KEY TERMS

---

science	pseudoexplanation
scientist	circular explanation or tautology
basic research	method of authority
applied research	rational method
confirmation bias	scientific method
pseudoscience	variable
scientific explanation	hypothesis
parsimonious explanation	deductive reasoning
commonsense explanations	pilot study
belief-based explanation	

## 2

## CHAPTER

# Developing and Evaluating Theories of Behavior

## CHAPTER OUTLINE

### What Is a Theory?

- Theory Versus Hypothesis
- Theory Versus Law
- Theory Versus Model
- Mechanistic Explanations Versus Functional Explanations

### Classifying Theories

- Is the Theory Quantitative or Qualitative?
- At What Level of Description Does the Theory Operate?
- What Is the Theory's Domain?

### Roles of Theory in Science

- Understanding
- Prediction
- Organizing and Interpreting Research Results
- Generating Research

### Characteristics of a Good Theory

- Ability to Account for Data
- Explanatory Relevance
- Testability
- Prediction of Novel Events
- Parsimony

### Strategies for Testing Theories

- Following a Confirmational Strategy
- Following a Disconfirmational Strategy
- Using Confirmational and Disconfirmational Strategies Together
- Using Strong Inference

### Theory-Driven Versus Data-Driven Research

### Summary

### Key Terms

As noted in Chapter 1, a major goal of any science is to develop valid, general explanations for the phenomena within its field of inquiry, and this is just as true of psychology as of any other science. A considerable portion of the research effort in psychology focuses on the development and testing of psychological theories: proposed explanations for observed psychological phenomena. Because so many research studies are designed at least partly for the purpose of testing and evaluating the merits of one or another theory, we thought it important for you to have a firm grasp of what theories are, how they are developed, and how to go about evaluating them against a number of criteria, before turning our attention to the “nuts and bolts” of the research process in the next chapter. We begin by defining “theory” and distinguishing it from some related terms.

## WHAT IS A THEORY?

In everyday discourse we tend to use the word *theory* rather loosely, to describe everything from well-tested explanations for some event to simple guesses that seem consistent with whatever information we happen to possess. In science, however, the term refers to something more specific. A **scientific theory** is one that goes beyond the level of a simple hypothesis, deals with potentially verifiable phenomena, and is highly ordered and structured. This discussion adopts and extends the definition of theory provided by Martin (1985): A *theory* is a partially verified statement of a scientific relationship that cannot be directly observed. If the theory is stated formally, this statement consists of a set of interrelated propositions (and corollaries to those propositions) that attempt to specify the relationship between a variable (or set of variables) and some behavior. Not all scientific theories are expressed this way, but most could be.

A good example of a psychological theory with a clearly defined set of propositions and corollaries is *equity theory* (Walster, Walster, &



**TABLE 2-1 Propositions and Corollaries of Equity Theory**

1. In an interpersonal relationship, a person will try to maximize his or her outcomes (where outcome = rewards – costs).  
*Corollary:* As long as a person believes that he or she can maximize outcomes by behaving equitably, he or she will. If a person believes that inequitable behavior is more likely to maximize outcomes, inequitable behavior will be used.
- 2a. By developing systems whereby resources can be equitably distributed among members, groups can maximize the probability of equitable behavior among their members.
- 2b. A group will reward members who behave equitably toward others and punish those who do not.
3. Inequitable relationships are stressful for those within them. The greater the inequity, the greater the distress.
4. A person in an inequitable relationship will take steps to reduce the distress aroused by restoring equity. The more distress felt, the harder the person will try to restore equity.

SOURCE: Based on Walster, Walster, and Berscheid, 1978.

Berscheid, 1978). Equity theory was developed to explain how individuals behave when placed in an interpersonal exchange situation, such as employer–employee relations or friendships. Table 2-1 presents the major propositions of equity theory and one corollary (other corollaries are outlined by Walster et al.). Notice that the first set of propositions makes a general statement about how interpersonal exchanges are perceived. The later propositions and corollaries specify how a set of variables (such as inputs and outputs) should affect the perception of equity within a relationship.

A deeper exploration of the definition of *theory* shows that a scientific theory has several important characteristics. First, a scientific theory describes a scientific relationship—one inferred through observation and logic—that indicates how variables interact within the system to which the theory applies.

Second, the described relationship cannot be observed directly. Its existence must be *inferred* from the data. (If you could observe the relationship directly, there would be no need for a theory.) Third, the statement is only partially verified. This means that the theory has passed some tests but that not all relevant tests have been conducted.

Colloquial use of the term *theory* leads to confusion over what a theory really is. You can also find confusion within the scientific community over the term. Even in scientific writing, “theory,” “hypothesis,” “law,” and “model” are often used interchangeably. Nevertheless, these terms can be distinguished, as described in the next sections.

### Theory Versus Hypothesis

Students often confuse theory with hypothesis, and even professionals sometimes use these terms interchangeably. However, as usually defined, theories are more complex

than hypotheses. For example, if you observe that more crime occurs during the period of full moon than during other times of the month, you might hypothesize that the observed relationship is caused by the illumination that the moon provides for nighttime burglary. You could then test this hypothesis by comparing crime rates during periods of full moon that were clear with crime rates during periods of full moon that were cloudy.

In contrast to the simple one-variable account provided by this hypothesis, a *theory* would account for changes in crime rate by specifying the action and interaction of a *system* of variables. Because of the complexity of the system involved, no single observation could substantiate the theory in its entirety.

### Theory Versus Law

A theory that has been substantially verified is sometimes called a **law**. However, most laws do not derive from theories in this way. Laws are usually empirically verified, quantitative relationships between two or more variables and thus are not normally subject to the disconfirmation that theories are. For example, the *matching law* was originally proposed by Richard Herrnstein (1970) to describe how pigeons divide their keypecks between two response keys associated with two different variable-interval schedules of reinforcement. According to the matching law, the relative rate of responding on a key (the percentage of responses directed to that key per unit of time) will match the relative rate of reinforcement (the percentage of reinforcers delivered on the schedule associated with that key per unit of time). The matching law has been found to hold under a variety of conditions beyond those for which it was originally formulated and even has been shown to describe the proportion of shots taken by basketball players from beyond the three-point line (Vollmer & Bourret, 2000) and the relative ratio of passing plays to rushing plays in football (Reed, Critchfield, & Martins, 2006).

Sometimes laws idealize real-world relationships—for example, Boyle's law, which relates change in temperature to change in pressure of a confined ideal gas. Because there are no ideal gases, the relationship described by Boyle's law is not directly observable. However, as a description of the behavior of real gases, it holds well enough for most purposes. To an approximation, it represents a verified empirical relationship and is thus unlikely to be overthrown.

Such empirical laws are not highly verified theories. They are relationships that must be explained by theory. The matching law, for example, merely describes how behavior is allocated among alternatives; it does not *explain* why matching occurs. For that, you need a theory. To explain matching, Herrnstein and Prelec (1992) proposed that an individual repeatedly samples the ratio of responses to reinforcements associated with each option and responds by moving toward the more favorable alternative—a process they termed “melioration.” (Melioration theory is but one of several proposed theories of matching the scientific community is currently evaluating.)

### Theory Versus Model

Like *theory*, the term *model* can refer to a range of concepts. In some cases, it is simply used as a synonym for *theory*. However, in most cases **model** refers to a specific implementation of a more general theoretical view. For example, the Rescorla–Wagner model of classical conditioning formalizes a more general associative theory of conditioning

(Rescorla & Wagner, 1972). This model specifies how the associative strength of a conditional stimulus (CS) is to be calculated following each of a series of trials in which the CS is presented alone or in conjunction with other stimuli. Where the general associative theory simply states that the strength of the stimulus will increase each time that the CS is paired with an unconditional stimulus (US), the Rescorla–Wagner model supplies a set of assumptions that mathematically specifies how characteristics of the stimuli interact on each trial to produce observed changes in response strength.

Rescorla and Wagner (1972) made it clear when they presented their model that the assumptions were simply starting points. For example, they assumed that the associative strength of a compound stimulus (two or more stimuli presented together) would equal the sum of the strengths of the individual stimuli. If the learning curves resulting from this assumption proved not to fit the curves obtained from experiment, then the assumption would be modified.

Rescorla and Wagner (1972) could have chosen to try several rules for combining stimulus strengths. Each variation would represent a somewhat different model of classical conditioning although all would derive from a common associative view of the conditioning process.

In a related sense, a model can represent an application of a general theory to a specific situation. In the case of the Rescorla–Wagner model, the assumptions of the model can be applied to generate predictions for reinforcement of a simple CS, for reinforcement of a compound CS, for inhibitory conditioning, for extinction, and for discrimination learning (to name a few). The assumptions of the model remain the same across all these cases, but the set of equations required in order to make the predictions changes from case to case. You might then say that each set of equations represents a different model: a model of simple conditioning, a model of compound conditioning, a model of differential conditioning, and so on. However, all the models would share the same assumptions of the Rescorla–Wagner model of conditioning.

**Computer Modeling** Theories in psychology most commonly take the form of a set of verbal statements that describe their basic assumptions and the ways in which the various entities of the theory interact to produce behavior. Unfortunately, predictions based on such theories must be derived by verbally tracing a chain of events from a set of initial conditions to the ultimate result, a difficult process to carry out successfully in an even moderately complex theory and one that may be impossible if the theory involves entities that mutually influence one another. Because of these difficulties, scientists may at times disagree about what a given theory predicts under given circumstances.

One way to avoid such problems is to cast specific implementations of the theory in the form of a computer model. A *computer model* is a set of program statements that define the variables to be considered and the ways in which their values will change over the course of time or trials. The process of creating a model forces you to be specific: to state precisely what variables are involved, what their initial values or states will be, and how the variables will interact. Developing a computer model offers several advantages:

1. The attempt to build a computer model may reveal inconsistencies, unspoken assumptions, or other defects in the theory and thus can help bring to light problems in the theory that otherwise might go unnoticed.

2. Having a computer model eliminates ambiguity; you can determine exactly what the model assumes by examining the code.
3. A properly implemented computer model will show what is to be expected under specified conditions. These predictions may be difficult or impossible to derive correctly by verbally tracing out the implications of the theory.
4. The behavior of the model under simulated conditions can be compared with the behavior of real people or animals under actual conditions to determine whether the model behaves realistically. Discrepancies reveal where the model has problems and may suggest how the model can be improved.
5. Competing theories can be evaluated by building computer models based on each and then determining which model does a better job of accounting for observed phenomena.

An interesting example of computer modeling is provided by Josef Nerb and Hans Spada (2001). Nerb and Spada were interested in explaining the relationship between cognitive, emotional, and behavioral responses to environmental disasters. More specifically, they were interested in investigating the relationship between media portrayals of single-event environmental disasters and cognitive, emotional, and behavioral responses to them. They developed a computer model called “Intuitive Thinking in Environmental Risk Appraisal,” or ITERA for short. The ITERA model was designed to make predictions about the cognitive appraisals made about environmental disasters as well as the emotions and behavioral tendencies generated in response to the disasters. By inputting data into the model relating to several variables, one can make predictions about cognitive, emotional, and behavioral outcomes. Those predictions could then be tested empirically to verify the validity of the computer model.

Let’s see how the model works. Nerb and Spada (2001) extracted crucial pieces of information from media reports of environmental disasters that related to elements of the ITERA model. This information was systematically varied and entered as variables into the model. For example, one piece of information referred to damage done by the disaster. This was entered into the model as “given,” “not given,” or “unknown.” The same protocol was followed for other variables (e.g., extent to which the events surrounding the disaster were controllable). By systematically entering one or more variables into the model, predictions about cognitive, emotional, and behavioral responses can be made. For example, the computer model predicts that if controllability information is entered indicating that the events leading to the disaster were controllable, anger should be a stronger emotion than sadness, and boycott behavior should be preferred over providing help. In contrast, the model predicts that if the disaster was uncontrollable, the dominant emotion should be sadness, and the dominant behavioral tendency would be to offer help.

Nerb and Spada (2001) tested this prediction in an experiment in which participants read a fictitious but realistic newspaper account of a tanker running aground in the North Sea, spilling oil into the sea. There were three versions of the newspaper article. In one version, participants were told that “the tanker did not fulfill safety guidelines, and the damage could have been avoided.” In a second version, participants read that “the tanker did fulfill safety guidelines, and the damage could not

have been avoided.” In the third condition, no information was provided about safety guidelines or whether the damage could have been avoided. The results from the experiment were compared to the predictions made by the ITERA computer model. Nerb and Spada found that the model correctly predicted emotional and behavioral outcomes for the controllable condition (safety guidelines not followed and the damage could have been avoided). Consistent with the model’s prediction, the dominant emotion reported by participants was anger, and the favored behavioral response was a boycott. However, the model did not correctly predict outcomes for an uncontrollable event (guidelines followed and damage was unavoidable). In this condition, sadness and helping did not dominate.

### Mechanistic Explanations Versus Functional Explanations

Theories provide explanations for observed phenomena, but not all explanations are alike. When evaluating a theory, you should carefully note whether the explanations provided are mechanistic or functional. A **mechanistic explanation** describes the mechanism (physical components) and the chain of cause and effect through which conditions act on the mechanism to produce its behavior; it describes *how* something works. In contrast, a **functional explanation** describes an attribute of something (such as physical attractiveness) in terms of its function—that is, what it does (e.g., in women, beauty signals reproductive health, according to evolutionary psychologists); it describes *why* the attribute or system exists. To clarify this distinction, consider the notion of motivated reasoning, which involves goals and motives influencing one’s reasoning process (Kunda, 1990). Kunda describes the mechanisms involved in motivated reasoning (e.g., optimistic reasoning) by pointing to the idea that individuals using motivated reasoning come up with a set of reasonable justifications for their conclusions. So a woman may convince herself that the chance of surviving breast cancer is excellent. However, she also must develop justifications for her optimism (e.g., that she is a strong person, that she will adhere to her treatment schedule rigorously). Contrast this with the more functional explanation for optimism provided by Shelly Taylor (1989). Taylor explains optimism in terms of its function of helping a person get better faster.

Mechanistic explanations tell you how a system works without necessarily telling you why it does what it does; functional explanations refer to the purpose or goal of a given attribute or system without describing how those purposes or goals are achieved. A full understanding requires both types of explanation.

Although you can usually determine a mechanism’s function once you know how it works, the converse is not true. Knowing what a system does gives only hints as to the underlying mechanism through which its functions are carried out. Consider, for example, the buttons on your television’s remote control. You can quickly determine their functions by trying them out—this one turns on the power, that one changes the volume, the next one changes the channel. However, without some knowledge of electronics, you may have no idea whatever how this or that button accomplishes its function, and even with that knowledge, there may be dozens of different circuits (mechanisms) that can do the job. Knowing *what* a button does in no way tells you *how* it does it.

Given the choice between a mechanistic explanation and a functional one, you should prefer the mechanistic one. Unfortunately, arriving at the correct mechanism underlying a given bit of human or animal behavior often may not be possible given our current understanding of the brain. For example, we currently have no firm idea how memories are stored in the brain and subsequently accessed (although there has been plenty of speculation). Yet we do have a fair understanding of many functional properties of the brain mechanism or mechanisms involved. Given this knowledge, it is possible to construct a theory of, say, choice among alternatives not currently present that simply *assumes* a memory with certain properties without getting into the details of mechanism.

### QUESTIONS TO PONDER

---

1. What is the definition of a scientific theory?
2. How does a theory differ from a hypothesis, a law, and a model?
3. What is a computer model, and what are the advantages of designing one?
4. How do mechanistic and functional theories differ? Which type is better, and why?

### CLASSIFYING THEORIES

---

Theories can be classified along several dimensions. Three important ones are (1) quantitative or qualitative aspect, (2) level of description, and (3) scope (or domain) of the theory. In light of these distinctions, we've organized our discussion by posing three questions that you can ask about any theory:

1. Is the theory quantitative or qualitative?
2. At what level of description does the theory operate?
3. What is the theory's domain?

#### Is the Theory Quantitative or Qualitative?

The first dimension along which a theory can be classified is whether the theory is quantitative or qualitative. Here we describe the characteristics of each type.

**Quantitative Theory** A **quantitative theory** defines the relationships between its variables and constants in a set of mathematical formulas. Given specific numerical inputs, the quantitative theory generates specific numerical outputs. The relationships thus described then can be tested by setting up the specified conditions and observing whether the outputs take on the specified values (within the error of measurement).

A good example of a quantitative theory in psychology is information integration theory developed by Norman Anderson (1968). Anderson's theory attempts to

explain how diverse sources of information are integrated into an overall impression. The theory proposes that each item of information used in an impression formation task is assigned both a weight and scale value. The weights and scale values are then combined according to the following formula:

$$J = \frac{\sum(w_i s_i)}{\sum w_i}$$

where  $w_i$  is the weight assigned to each item of information and  $s_i$  is the scale value assigned to each item of information. According to this theory, your final judgment ( $J$ ) about a stimulus (e.g., whether you describe a person as warm or cold, caring or uncaring, honest or dishonest) will be the result of a mathematical combination of the weights and scale values assigned to each piece of information.

**Qualitative Theory** A **qualitative theory** is any theory that is not quantitative. Qualitative theories tend to be stated in verbal rather than mathematical terms. These theories state which variables are important and, loosely, how those variables interact.

The relationships described by qualitative theories may be quantitative, but if so, the quantities will be measured on no higher than an ordinal scale (as rankings, such as predicting that anxiety will increase, without specifying by how much). For example, a theory of drug addiction may state that craving for the drug will increase with the time since the last administration and that this craving will be intensified by emotional stress. Note that the predictions of the theory specify only ordinal relationships. They state that craving will be greater under some conditions than under others, but they do not state by how much.

A good example of a qualitative theory in psychology is a theory of language acquisition by Noam Chomsky (1965). This theory states that a child acquires language by analyzing the language that he or she hears. The language heard by the child is processed, according to Chomsky, and the rules of language are extracted. The child then formulates hypotheses about how language works and tests those hypotheses against reality. No attempt is made in the theory to quantify the parameters of language acquisition. Instead, the theory specifies verbally the important variables that contribute to language acquisition.

### At What Level of Description Does the Theory Operate?

The second dimension along which theories may be categorized is according to the level of description that the theory provides. Two goals of science are to describe and explain phenomena within its domain. A theory may address itself to the first goal (description), whereas another may address itself to the second (explanation). So some theories are primarily designed to describe a phenomenon whereas others attempt to explain relationships among variables that control a phenomenon. The following sections differentiate theories that deal with phenomena at different levels: descriptive, analogical, and fundamental.

**Descriptive Theories** At the lowest level, a theory may simply describe how certain variables are related without providing an explanation for that relationship. A theory that merely describes a relationship is termed a **descriptive theory**.

An example of a descriptive theory is Wilhelm Wundt's systematic theory of the structure of consciousness. Wundt, as you probably know, is credited with being the founder of scientific psychology. His empirical and theoretical work centered on describing the structure of consciousness. Wundt (1897) maintained that consciousness is made up of *psychical elements* (sensations, feelings, and volition). He stated that all examples of consciousness are made up of these three basic building blocks. When the psychical elements combined, they formed *psychical compounds*. Wundt focused on describing the structure of consciousness and how complex conscious events could be broken down into their component parts.

Most descriptive theories are simply proposed generalizations from observation. For example, arousal theory states that task performance increases with arousal up to some optimal arousal value and then deteriorates with further increases in arousal. The proposed relationship thus follows an inverted U-shaped function. Arousal and task performance are both classes of variables that can be operationally defined a number of ways. Arousal and task performance are general concepts rather than specific variables. The proposed relationship is thus not directly observable but must be inferred from observing many specific variables representative of each concept. Note also that the theory describes the relationship but offers no real explanation for it.

Descriptive theories provide only the weakest form of explanation. If you discover that being in elevators makes you nervous, then you could explain your current high level of anxiety by noting that you are standing in an elevator. But such an explanation does not tell you *why* elevators have the effect on you that they do.

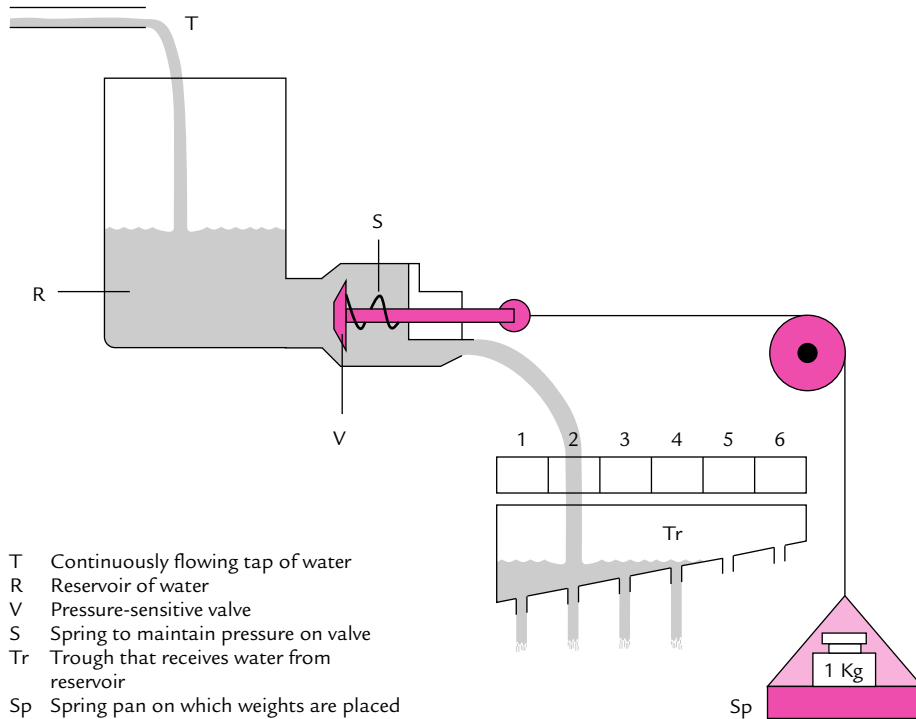
**Analogical Theories** At the next level is an **analogical theory**, which explains a relationship through analogy. Such theories borrow from well-understood models (usually of physical systems) by suggesting that the system to be explained behaves in a fashion similar to that described by the well-understood model.

To develop an analogical theory, you equate each variable in the physical system with a variable in the behavioral system to be modeled. You then plug in values for the new variables and apply the rules of the original theory in order to generate predictions.

An example of an analogical theory was provided by Konrad Lorenz (1950). Lorenz wanted to explain some relationships that he had observed between the occurrence of a specific behavioral pattern (called a *fixed-action pattern*, or FAP), a triggering stimulus (called a *sign* or *releaser stimulus*), and the time since the last occurrence of the FAP. For example, chickens scan the ground and then direct pecks at any seeds they find there. Here the visual characteristics of the seeds act as a releaser stimulus, and the directed pecking at the seeds is the FAP. Lorenz had observed that the FAP could be elicited more easily by a sign stimulus as the time increased since the last appearance of the FAP. In fact, with enough time, the behavior became so primed that it sometimes occurred in the absence of any identifiable sign stimulus. However, if the behavior had just occurred, the sign stimulus usually could not elicit the FAP again.

Let's return to the chicken example. Chickens at first peck only at seeds. With increasing hunger, however, they begin to peck at pencil marks on a paper and other such stimuli that only remotely resemble seeds. With further deprivation, they even peck at a blank paper.





**FIGURE 2-1** Lorenz's hydraulic model of motivation

SOURCE: Lorenz, 1950; legend adapted from Dewsbury, 1978; reprinted with permission.

To explain this relationship, Lorenz imagined that the motivation to perform the FAP was like the pressure of water at the bottom of a tank that was being continuously filled (see Figure 2-1). As time went on, the water in the tank became deeper and the pressure greater. Lorenz pictured a pressure-sensitive valve at the bottom of the tank. This valve could be opened by depressing a lever, but the pressure required to open it became less as the pressure inside the tank rose. In Lorenz's conception, the lever was normally "pressed" by the appearance of the sign stimulus.

Notice the analogies in Lorenz's model. Motivation to perform the FAP is analogous to water pressure. Engaging in the FAP is analogous to water rushing out the open valve. And perception of the sign stimulus is analogous to pressing the lever to open the valve.

Now put the model into action. Motivation to perform the FAP builds as time passes (the tank fills). If a sign stimulus appears after the tank has partially filled, the valve opens and the FAP occurs. However, if the sign stimulus does not occur for a long time, the tank overfills and the pressure triggers the valve to open spontaneously (the FAP occurs without the sign stimulus). Finally, if the FAP has just occurred (the valve has just opened), there is no motivation to perform the FAP (the tank is empty), and the sign stimulus is ineffective. The model thus nicely accounts for the observed facts.

Lorenz's hydraulic model of motivation eventually gave way to more sophisticated theories when new data revealed its limitations. In general, analogical theories can be pushed only so far. At some point, the analogy breaks down. After all, motivation is *not* quite the same thing as water pressure in a tank and may vary in ways quite unexpected for water pressure. Nevertheless, analogical theories can provide conceptual organization for the data and may predict relationships that otherwise would be unexpected.

**Fundamental Theories** At the highest level are theories created to explain phenomena within a particular area of research. These theories do not depend on analogy to provide their basic structures. Instead, they propose a new structure that directly relates the variables and constants of the system. This structure includes entities and processes not directly observable but invented to account for the observed relationships. Thus, these entities and processes go beyond descriptive theories, which simply describe relationships among observable variables. Because these theories have no accepted name, we'll call this type of theory a **fundamental theory** to distinguish them from the more superficial descriptive and analogical types. Such theories seek to model an underlying reality that produces the observed relationships among variables. In this sense, they propose a more fundamental description of reality than the analogical theory.

Although psychological theories abound, fundamental theories are disturbingly rare in psychology. Part of the reason for this rarity is that psychology is still a relatively new science, but this is probably only a small part. Mostly this rarity is because of the complexity of the system being studied and because of the extreme difficulty in controlling the relevant variables well enough to clearly reveal the true relationships among them (or even to measure them properly). The physicist can expect every electron to behave exactly like every other. The psychologist cannot even hope that his or her subjects will be this interchangeable. Nevertheless, some attempts at fundamental theorizing have been made. One of the most famous fundamental theories is *cognitive dissonance theory* proposed by Festinger (1957).

According to the theory, dissonance is the fundamental process in cognitive dissonance theory. Whenever two (or more) attitudes or behaviors are inconsistent, a negative psychological state called *cognitive dissonance* is aroused. The arousal of dissonance motivates the individual to reduce dissonance. This can be done by changing behavior or by changing attitudes. Festinger's theory thus described how dissonance leads to behavioral or attitude change.

Another example of fundamental theory in psychology is the Scalar Expectancy Theory (SET) proposed by John Gibbon (1977) to account for the patterns of responding that develop under various schedules of reinforcement. The central idea of Gibbon's theory is that well-trained subjects are able to estimate time to reinforcement by means of a "scalar timing" process. With scalar timing, the subject can adjust to changes in the time constant of a schedule by simply rescaling the estimated time distribution to fit the new constant. Estimates of time to reinforcement (together with the size and attractiveness of the reinforcer) determine the "expectancy" of reward, which in turn determines the probability of a response through a well-defined mechanism. Gibbon described how the assumption of scalar timing produces a better

fit to data from a variety of paradigms than do other assumptions (such as timing based on a Poisson process).

### What Is the Theory's Domain?

The third dimension along which theories differ is **domain**, or *scope*. This dimension concerns the range of situations to which the theory may be legitimately applied. A theory with a wide scope can be applied to a wider range of situations than can a theory with a more limited scope.

Gibbon's (1977) Scalar Expectancy Theory is an example of a theory with a relatively limited scope. It provided an explanation for behavioral patterns that emerge under a wide variety of reinforcement schedules, but it did not attempt to account for other factors that could affect behavior. Cognitive consistency theory, such as Festinger's (1957) theory of cognitive dissonance, is an example of a theory with a wider scope. It has been applied beyond attitude change (for which it was developed) to help explain motivational processes in other contexts.

The chances of dealing adequately with a range of phenomena are better for a small area of behavior than they are for a large area. On the negative side, however, concepts invented to deal with one area may have no relationship to those invented to deal with others, even though the behaviors may be mediated by partly overlapping (or even identical) mechanisms.

## QUESTIONS TO PONDER

---

1. What are the defining characteristics of quantitative and qualitative theories?
2. What is a descriptive theory?
3. What is an analogical theory?
4. What is a fundamental theory?
5. How do descriptive, analogical, and fundamental theories differ? Which is preferred and why?

## ROLES OF THEORY IN SCIENCE

---

Theories have several roles to play in science. These roles include providing an understanding of the phenomena for which they account, providing a basis for prediction, and guiding the direction of research.

### Understanding

At the highest level, theories represent a particular way to understand the phenomena with which they deal. To the degree that a theory models an underlying reality, this understanding can be deep and powerful. For example, Jean Piaget's (1952) theory of development provided a deep insight into the thought processes of children

and helped us better understand how these processes change with age and experience. Piaget provided a broad description of the behaviors that are characteristic of children at various ages. Within the theory, he also proposed mechanisms (organization, adaptation, and equilibration) to explain how development takes place.

### Prediction

Even when theories do not provide a fundamental insight into the mechanisms of a behaving system (as descriptive theories do not), they at least can provide a way to predict the behavior of the system under different values of its controlling variables. The descriptive theory will specify which variables need to be considered and how they interact to determine the behavior to be explained. If it is a good theory, the predictions will match the empirical outcome with a reasonable degree of precision.

A good example of how a theory can generate testable predictions comes from *social impact theory* proposed by Bibb Latané (1981). Social impact theory is intended to explain the process of social influence (e.g., conformity and obedience). According to the theory, the amount of influence obtained is dependent upon the interaction of three factors: the strength of an influence source ( $S$ ), the immediacy of an influence source ( $I$ ), and the number of influence sources ( $N$ ). The relationship between influence and these three variables is summed up with this simple formula:

$$\text{Influence} = \text{a function of } (S \times I \times N)$$

One prediction made by the theory is that the relationship between the number of sources and the amount of influence obtained is nonlinear. That is, after a certain number of sources, influence should not increase significantly and should “level off.” The prediction made from social impact theory is consistent with the results obtained from empirical research findings on the relationship between the size of a majority and conformity.

### Organizing and Interpreting Research Results

A theory can provide a sound framework for organizing and interpreting research results. For example, the results of an experiment designed to test Piaget’s theory will be organized within the existing structure of confirmatory and disconfirmatory results. This organization is preferable to having a loose conglomeration of results on a topic.

In addition to being organized by theory, research results can be interpreted in the light of a theory. This is true even if your research was not specifically designed to test a particular theory. For example, results of a study of decision making may be interpreted in the light of cognitive dissonance theory even though you did not specifically set out to test dissonance theory.

### Generating Research

Finally, theories are valuable because they often provide ideas for new research. This is known as the *heuristic value* of a theory. The heuristic value of a theory is often independent of its validity. A theory can have heuristic value even when it is not supported by subsequent empirical research. Such a theory may implicate

certain variables in a particular phenomenon, variables that had not been previously suspected of being important. Researchers may then design experiments or collect observations to examine the role of these variables. Often such variables turn out to be significant although the theory that emphasized them eventually may be proved wrong.

A theory specifies the variables that need to be examined and the conditions under which they are to be observed and may even state how they are to be measured. It provides a framework within which certain research questions make sense and others become irrelevant or even nonsensical. Franz Gall's phrenology provides an example of how a theory guides research and determines which questions will be considered important. Gall was a 19th-century surgeon who was convinced that a person's abilities, traits, and personality were determined by specific areas of the cerebral cortex. If a part of the brain were highly developed, Gall believed, a person would have a higher degree of the particular trait or ability associated with that area than if that same part of the brain were less highly developed. In addition, Gall reasoned, the more highly developed area would require more volume of cortex. Consequently, the part of the skull covering this area would bulge outward and create a "bump" on the person's head (Fancher, 1979).

In the context of Gall's theory, the important research problems were to identify which parts of the cortex represented which traits or abilities and to relate individual differences in the topography of the skull (its characteristic bumps and valleys) to personality variables. Researchers developed special instruments to measure the skull and devoted thousands of hours to making measurements and collecting profiles of mental abilities and personality traits.

Phrenology never gained acceptance within the scientific community and was severely damaged by evidence (provided by Pierre Flourens and other of Gall's contemporaries) showing rather conclusively that at least some of the brain areas identified as the seat of a particular trait had entirely different functions (Fancher, 1979). With the discrediting of phrenology, interest in measuring the skull and in correlating these measurements with traits and abilities went with it.

Phrenology provided a framework for research within which certain problems and questions became important. When this view was displaced, much of the work conducted under it became irrelevant. This loss of relevance is a serious concern. If data collected under a particular theory become worthless when the theory dies, then researchers working within a particular framework face the possibility that the gold they mine will turn to drizzle in the future.

This possibility has led some researchers to suggest that perhaps theories should be avoided, at least in the early stages of research. Speaking at a time when the Hull-Spence learning theory was still a force within the psychology of learning, B. F. Skinner (1949) asked in his presidential address to the Midwestern Psychological Association, "Are theories of learning necessary?" In this address, Skinner disputed the claim that theories are necessary to organize and guide research. Research should be guided, Skinner said, not by theory but by the search for functional relationships and for orderly changes in data that follow the manipulation of effective independent variables. Such clearly established relationships have enduring value. These relationships become the data with which any adequate theory must deal.

Skinner's point has a great deal of merit, and it is discussed later in this chapter. However, for now you should not get the idea that theory is useless or, even worse, wasteful. Even theories that are eventually overthrown do provide a standard against which to judge new developments. New developments that do not fit the existing theory become anomalies, and anomalies generate further research in an effort to show that they result from measurement error or some other problem unrelated to the content of the theory. The accumulation of serious anomalies can destroy a theory. In the process, however, the intense focus on the problem areas may bring new insights and rapid progress within the field. Because anomalies are unexpected findings, they exist only in the context of expectation—expectation provided by theory. Thus, even in failing, a theory can have heuristic value.

## CHARACTERISTICS OF A GOOD THEORY

---

In the history of psychology, many theories have been advanced to explain behavioral phenomena. Some of these theories have stood the test of time, whereas others have fallen by the wayside. Whether or not a theory endures depends on several factors, including the following.

### Ability to Account for Data

To be of any value, a theory must account for most of the existing data within its domain. Note that the amount of data is “most” rather than “all” because at least some of the data may in fact be unreliable. A theory can be excused for failing to account for erroneous data. However, a theory that fails to account for well-established facts within its domain is in serious trouble. The phrase “within its domain” is crucial. If the theory is designed to explain the habituation of responses, it can hardly be criticized for its failure to account for schizophrenia. Such an account clearly would be beyond the scope of the theory.

### Explanatory Relevance

A theory also must meet the criterion of *explanatory relevance* (Hempel, 1966). That is, the explanation for a phenomenon provided by a theory must offer good grounds for believing that the phenomenon would occur under the specified conditions. If a theory meets this criterion, you should find yourself saying, “Ah, but of course! That was indeed to be expected under the circumstances!” (Hempel, 1966). If someone were to suggest that the rough sleep you had last night was caused by the color of your socks, you would probably reject this theory on the grounds that it lacks explanatory relevance. There is simply no good reason to believe that wearing a particular color of socks would affect your sleep. To be adequate, the theory must define some logical link between socks and sleep.

### Testability

Another condition that a good theory must meet is *testability*. A theory is testable if it is capable of failing some empirical test. That is, the theory specifies outcomes

under particular conditions, and if these outcomes do not occur, then the theory is rejected.

The criterion of testability is a major problem for many aspects of Freud's psychodynamic theory of personality. Freud's theory provides explanations for a number of personality traits and disorders, but it is too complex and loosely specified to make specific, testable predictions. For example, if a person is observed to be stingy and obstinate, Freudian theory points to unsuccessful resolution of the anal stage of psychosexual development. Yet diametrically opposite traits also can be accounted for with the same explanation. There is no mechanism within the theory to specify which will develop in any particular case. When a theory can provide a seemingly reasonable explanation no matter what the outcome of an observation, you are probably dealing with an untestable theory.

### Prediction of Novel Events

A good theory should predict new phenomena. Within its domain, a good theory should predict phenomena beyond those for which the theory was originally designed. Strictly speaking, such predicted phenomena do not have to be new in the sense of not yet observed. Rather, they must be new in the sense that they were not taken into account in the formulation of the theory.

As an example, consider the Rescorla–Wagner model of classical conditioning we described previously in the chapter. The model predicts that when two fully conditioned stimuli are presented together, the resulting compound CS initially will evoke an even stronger response than either single stimulus presented alone, a phenomenon now called “overexpectation.” Furthermore, the model predicted that further pairings of the compound CS with the unconditioned stimulus would cause the conditioned response to weaken—a surprising result given that such a “reinforcement” process normally would be expected to strengthen, not weaken, the response. Appropriate tests confirmed both of these predictions.

### Parsimony

The medieval English philosopher William of Ockham popularized an important principle stated by Aristotle. Aristotle's principle states, “Entities must not be multiplied beyond what is necessary” (Occam's Razor, n.d.). Ockham's refinement of this principle is now called *Occam's Razor* and states that a problem should be stated in the simplest possible terms and explained with the fewest postulates possible. Today we know this as the *law of parsimony*. Simply put, a theory should account for phenomena within its domain in the simplest terms possible and with the fewest assumptions. If there are two competing theories concerning a behavior, the one that explains the behavior in the simplest terms is preferred under the law of parsimony.

Many theories in psychology fit this requirement very well. Modern theories of memory, attribution processes, development, and motivation all adhere to this principle. However, the history of science in general, and psychology in particular, is littered with theories that were crushed under their own weight of complexity.

For example, the collapse of interest in the Hull–Spence model of learning occurred primarily because the theory had been modified so many times to account

for anomalous data (and had in the process gained so many ad hoc assumptions) that it was no longer parsimonious. Researchers could not bring themselves to believe that learning could be *that* complicated.

## QUESTIONS TO PONDER

---

1. What roles do theories play in science? Describe each role in detail.
2. What are the defining characteristics of a “good” theory? Describe each characteristic in detail.

## STRATEGIES FOR TESTING THEORIES

---

A major theme developed in the preceding sections is that a good scientific theory must be testable with empirical methods. In fact, the final step in the business of theory construction is to subject the propositions of your theory to rigorous empirical scrutiny.

### Following a Confirmational Strategy

A theory is usually tested by identifying implications of the theory for a specific situation not yet examined and then setting up the situation and observing whether the predicted effects occur. If the predicted effects are observed, the theory is said to be supported by the results, and your confidence in the theory increases. If the predicted effects do not occur, then the theory is not supported, and your confidence in it weakens.

When you test the implications of a theory in this way, you are following what is called a **confirmational strategy** (i.e., a strategy of looking for confirmation of the theory’s predictions). A positive outcome supports the theory.

Looking for confirmation is an important part of theory testing, but it does have an important limitation. Although the theory must find confirmation if it is to survive (too many failures would kill it), you can find confirmation until doomsday, and the theory may still be wrong.

Spurious confirmations are particularly likely to happen when the prediction only loosely specifies an outcome. For example, in an experiment with two groups, if a theory predicts that Group A will score higher on the dependent measure than Group B, only three outcomes are possible at this level of precision: A may be greater than B, B may be greater than A, or A and B may be equal. Thus, the theory has about a 1-in-3 chance of being supported by a lucky coincidence.

Such coincidental support becomes less likely as the predictions of the theory become more precise. For example, if the theory predicts that Group A will score 25, plus or minus 2, points higher than Group B, it is fairly unlikely that a difference in this direction and within this range will occur by coincidence. Because of this relationship, confirmation of a theory’s predictions has a much greater impact on your confidence in the theory when the predictions are precisely stated than when they are loosely stated.



### Following a Disconfirmational Strategy

Even when a theory's predictions are relatively precise, many alternative theories could potentially be constructed that would make the same predictions within the stated margin of error. Because of this fact, following a confirmational strategy is not enough. To test a theory requires more than simply finding out if its predictions are confirmed. You also must determine whether outcomes *not* expected, according to the theory, do or do not occur.

This strategy follows this form: If A is true (the theory is correct), then B will not be true (a certain outcome will not occur); thus, if B is true (the outcome does happen), then A is false (the theory is erroneous). Because a positive result will *disconfirm* (rather than confirm) the prediction, this way of testing a theory is called a **disconfirmational strategy**.

### Using Confirmational and Disconfirmational Strategies Together

Adequately testing a theory requires using both confirmational and disconfirmational strategies. Usually, you will pursue a confirmational strategy when a theory is fresh and relatively untested. The object during this phase of testing is to determine whether the theory can predict or explain the phenomena within its domain with reasonable precision.

If the theory survives these tests, you will eventually want to pursue a disconfirmational strategy. The objective during this phase of testing is to determine whether outcomes that are unexpected from the point of view of the theory nevertheless happen. If unexpected outcomes do occur, it means that the theory is, at best, incomplete. It will have to be developed further so that it can account for the previously unexpected outcome, or it will have to be replaced by a better theory.

### Using Strong Inference

The usual picture of progress in science is that theories are subjected to testing and then gradually modified as the need arises. The theory evolves through a succession of tests and modifications until it can handle all extant data with a high degree of precision. This view of science has been challenged by Thomas Kuhn (1970). According to Kuhn, the history of science reveals that most theories continue to be defended and elaborated by their supporters even after convincing evidence to the contrary has been amassed. People who have spent their professional careers developing a theoretical view have too much invested to give up the view. When a more adequate view appears, the supporters of the old view find ways to rationalize the failures of their view and the successes of the new one. Kuhn concluded that the new view takes hold only after the supporters of the old view actually die off or retire from the profession. Then a new generation of researchers without investment in either theory objectively evaluates the evidence and makes its choice.

Commitment to a theoretical position well beyond the point at which it is objectively no longer viable is wasteful of time, money, and talent. Years may be spent evaluating and defending a view, with nothing to show for the investment. According to John Platt (1964), this trap can be avoided. Platt stated that the way to progress

in science is to develop several alternative explanations for a phenomenon. Each of these alternatives should give rise to testable predictions. To test the alternatives, you try to devise experiments whose outcomes can support only one or a few alternatives while ruling out the others.

When the initial experiment has been conducted, some of the alternatives will have been ruled out. You then design the next experiment to decide among the remaining alternatives. You continue this process until only one alternative remains. Platt (1964) called this process **strong inference**.

Strong inference can work only if the alternative explanations generate well-defined predictions. In biochemistry (the field that Platt, 1964, uses to exemplify the method), strong inference is a viable procedure because of the degree of control that scientists have over variables and the precision of their measures. The procedure tends to break down when the necessary degree of control is absent (so that the data become equivocal) or when the alternatives do not specify outcomes with sufficient precision to discriminate them. Unfortunately, in most areas of psychology, the degree of control is not sufficient, and the theories (usually loosely stated verbalizations) generally predict little more than the fact that one group mean will be different from another.

Nevertheless, Platt's (1964) approach can often be applied to test specific assumptions within the context of a particular view. In this case, applying strong inference means developing alternative models of the theory and then identifying areas in which clear differences emerge in predicted outcomes. The appropriate test then can be performed to decide which assumptions to discard and which to submit to further testing.

If several theories have been applied to the same set of phenomena and if these theories have been specified in sufficient detail to make predictions possible, you also may be able to use the method of strong inference if the theories make opposing predictions for a particular situation. The outcome of the experiment, if it is clear, will lend support to one or more of the theories while damaging others. This procedure is much more efficient than separately testing each theory, and you should adopt it wherever possible.

You now should have clear ideas about how to recognize, develop, and test adequate theories. However, an important question remains to be addressed: Should research be directed primarily toward testing theories or toward discovering empirical relationships?

## QUESTIONS TO PONDER

---

1. What is meant by confirmation and disconfirmation of a theory?
2. How are theories tested?
3. What is the difference between a confirmational and a disconfirmational strategy? How are they used to test a theory?
4. What is strong interference, and how is it used to test a theory?

## THEORY-DRIVEN VERSUS DATA-DRIVEN RESEARCH

---

At one time in the not-too-distant history of psychology, research efforts in one field centered on developing a theory of learning. This theory would organize and explain data obtained from many experiments involving white laboratory rats running down straight alleys, learning discrimination tasks, and finding their ways through mazes. Ultimately, this was to be a mathematical theory, complete with equations relating theoretical entities to each other and to observable variables.

The task of developing such a theory was taken up by Clark Hull at Iowa State University and by Hull's student, Kenneth Spence. Hull's approach to theory development was to follow the "hypothetico-deductive method," which consisted of adopting specific assumptions about the processes involved in learning, deriving predictions, submitting these predictions to experimental test, and then (as required) modifying one or more assumptions in the light of new evidence. Applied at a time when very few data were in fact available, the method was remarkably successful in producing an account that handled the relevant observations. This initial success galvanized researchers in the field, and soon it seemed that nearly everyone was conducting experiments to test the Hull–Spence theory.

The new data quickly revealed discrepancies between prediction and outcome. Some researchers, such as Edwin Tolman, rejected some of the key assumptions of the Hull–Spence theory and proposed alternative views. However, they were never able to develop their positions completely enough to provide a really viable theory of equivalent scope and testability. Besides, every time that Tolman and others would find an outcome incompatible with the Hull–Spence view, Hull and Spence would find a way to modify the theory in such a way that it would now account for the new data. The theory evolved with each new challenge.

These were exciting times for researchers in the field of learning. The development of a truly powerful, grand theory of learning seemed just around the corner. Then, gradually, things began to come apart. Hull died in 1952. Even before his death, discontent was beginning to set in, and even the continued efforts of Spence were not enough to hold researchers' interest in the theory.

Interest in the Hull–Spence theory collapsed for a number of reasons. Probably the most significant reason was that it had simply become too complex, with too many assumptions and too many variables whose values had to be extracted from the very data that the theory was meant to explain. Like the Ptolemaic theory of planetary motion, the system could predict nearly any observation (after the fact) once the right constants were plugged in—but it had lost much of its true predictive power, its parsimony, and its elegance.

With the loss of interest in the Hull–Spence theory went the relevance of much of the research that had been conducted to test it. Particularly vulnerable were those experiments that manipulated some set of variables in a complex fashion in order to check on some implication of the theory. These experiments demonstrated no clear functional relationship among simple variables, and the results were therefore of little interest except within the context of the theory. Viewed outside this context, the research seemed a waste of time and effort.

It was a tough lesson for many researchers. Much of the time and effort spent theorizing, tracing implications of the theory, developing experimental tests, and conducting observations was lost. This experience raises several questions concerning the use of theory in psychology. Should you attempt to develop theories? If you should develop theories, at what point should you begin? Should you focus your research efforts on testing the theories that you do develop?

The answer to the first question is definitely yes; you should attempt to develop theories. The history of science is littered with failed theories: the Ptolemaic system of astronomy, the phlogiston theory of heat, Gall's phrenology—the list goes on. In each case, much of the theorizing and testing became irrelevant when the theory was discarded. However, in each case, the attempt to grapple with the observations (particularly the anomalous ones) eventually led to the development of a more adequate theory. In this sense, the earlier efforts were not wasted.

Furthermore, it is the business of science to organize the available observations and to provide a framework within which the observations can be understood. At some point, theories must be developed if psychology is to progress.

The real question is not whether you should develop theories, but when. The major problem with the Hull–Spence theory is probably that it was premature. The attempt was made to develop a theory of broad scope before there was an adequate empirical database on which to formulate it. As a result, the requirements of the theory were not sufficiently constrained. The assumptions had to be repeatedly modified as new data became available, making some tests obsolete even before they could be published.

To avoid this problem, a theory that is more than a simple hypothesis should await the development of an adequate observational base. A sufficient number of well-established phenomena and functional relationships should be available to guide theory development and demonstrate the power of the resulting formulation.

The third question asked to what extent you should focus your research efforts on testing the theories that you do develop. There is no general agreement on the answer to this question. For one side of the issue, consider the letter written to *Science* by Bernard Forscher (1963) entitled “Chaos in the Brickyard.”

Forscher's (1963) letter presented an allegory in which scientists were compared to builders of brick edifices. The bricks were facts (observations), and the edifices were theories. According to Forscher's story, at one time the builders made their own bricks. This was a slow process, and the demand for bricks was always ahead of the supply. Still, the bricks were made to order, guided in their manufacture by a blueprint called a *theory* or *hypothesis*.

To speed the process, a new trade of brickmaking was developed, with the brickmakers producing bricks according to specifications given by the builders. With time, however, the brickmakers became obsessed with making bricks and began to create them without direction from the builders. When reminded that the goal was to create edifices, not bricks, the brickmakers replied that when enough bricks had been made, the builders could select the ones they needed.

Thus, it came to pass that the land was flooded with bricks. For the builders, constructing an edifice became impossible. They had to examine hundreds of bricks to find a suitable one, and it was difficult to find a clear spot of ground on which to build. Worst of all, little effort was made to maintain the distinction between an edifice and a pile of bricks.

Forscher's message was that experimentation conducted without the guidance of theory produces a significant amount of irrelevant information that is likely to obscure the important observations. From the infinite number of potential observations you could make, you need to select just those observations that will contribute most to progress in understanding. Theory provides one rationale for making that selection.

However, theory does not provide the only guide to choosing what observations to make. Observation also can be guided by the systematic exploration of functional relationships within a well-defined domain. This empirical approach was forcefully defended by B. F. Skinner in his 1949 address to the Midwestern Psychological Association.

Much of the research conducted in psychology has followed this program. A systematic study of memory by Ebbinghaus (1885/1964), and the work that followed it, provides a case in point. Ebbinghaus invented a nearly meaningless unit to memorize (the consonant–vowel–consonant, or CVC, trigram) and several methods to measure the strength of memory for the CVCs. He then systematically explored the effects of many variables in a series of parametric experiments. These variables included the amount of practice, spacing of practice, length of the retention interval, and serial position of the CVC within the list. The resulting functional relationships between these variables and retention were subsequently shown to be highly reliable phenomena.

The data from such observations provide the reliable phenomena that any subsequently developed theory must explain. As Skinner (1949) and others have indicated, these data stand independent of any particular theoretical view. Thus, if an experiment is designed to clearly illuminate simple functional relationships among variables—even when the experiment is conducted mainly for the purpose of testing theory—then the data will retain their value even if the theory is later discarded.

What conclusions can you draw from this discussion? First, the choice of observations to make can be guided both by theory and by a plan of systematic exploration. Second, guidance by theory is more likely to be of value when sufficient observations already have been conducted to construct a reasonably powerful theory. Third, even when theory testing is the major goal of the research, designing the study to illuminate simple functional relationships among the variables, if possible, ensures that the resulting observations will continue to have value beyond the usefulness of the theory.

Chapter 1 indicated that a science is an organized and systematic way of acquiring knowledge. Science is best advanced when results from research endeavors can be organized within some kind of framework. In many cases, results from both basic and applied research can be understood best when organized within a theory. Keep in mind, however, that not all research *must* be organized within a theoretical framework. Some purely applied research, for example, may best be organized with other research that also was geared toward the solution of a specific problem. Nevertheless, theory plays a central role in advancing science.

## QUESTIONS TO PONDER

---

1. How do theory-driven research and data-driven research differ?
2. What are the relative advantages and disadvantages of theory-driven and data-driven research?

## SUMMARY

---

A theory is a partially verified statement concerning the relationship among variables. A theory usually consists of a set of interrelated propositions and corollaries that specify how variables relate to the phenomena to be explained. *Hypothesis*, *law*, and *model* are all terms that are often used as synonyms for *theory*. There are, however, important differences among them. A hypothesis is a specific statement about a relationship that is subjected to direct empirical test. A law is a relationship that has received substantial support and is not usually subject to disconfirmation as theories are. A model is a specific implementation of a more general theoretical perspective. Models therefore usually have a more limited domain than do theories.

Computer models test the implications of a theory by encoding the theory as a series of program statements, supplying a set of initial conditions, and then observing how the model behaves. Such models remove ambiguity in the specific application of a theory and can reveal predictions of the theory that cannot be deduced by mere verbal reasoning. The behavior of the model under simulated conditions can be compared with the actual behavior of people or animals to determine whether the model behaves correctly, and alternative models can be compared to determine which does a better job of modeling actual behavior under given conditions.

Explanations provided by theories may be mechanistic or functional. Mechanistic explanations describe the physical components of a system and their connections (mechanism) whereas functional explanations describe only what the system does (function). Because function can be deduced from mechanism but mechanism cannot be uniquely deduced from function, you should prefer mechanistic theories over functional ones.

Theories vary along at least three dimensions. Some theories are quantitative in that they express relationships among variables in mathematical terms. Anderson's integration theory and the Rescorla–Wagner model of classical conditioning are examples of quantitative theories. Qualitative theories verbally express relationships among variables. No attempt is made to mathematically specify the nature of the relationships. Chomsky's theory of language acquisition is an example of a qualitative theory.

Theories also differ according to level of analysis. At the lowest level, descriptive theories simply seek to describe a phenomenon. At the next level, analogical theories try to explain phenomena by drawing parallels between known systems and the phenomenon of interest. At the highest level, fundamental theories represent new ways of explaining a phenomenon. These theories tend to provide a more fundamental look at a phenomenon than do descriptive or analogical theories. Finally, theories differ according to domain. A theory with a large domain accounts for more phenomena than does a theory with a more limited domain.

Theories play an important role in science. They help us to understand a phenomenon better, allow us to predict relationships, help us to organize and interpret our data, and, in many cases, help generate new research. This latter role is often independent of the correctness of the theory. Some theories, even though they are not correct, have led to important research and new discoveries that greatly advance science.

A theory must meet certain criteria before it can be accepted as a good theory. A theory must be able to account for most of the data within its domain. A theory

that does not do this is of little value. A good theory also must meet the criterion of explanatory relevance, which means that a theory must offer good grounds for believing that the phenomenon would occur under the specified conditions. An important criterion that any good theory must meet is that the theory be testable. The propositions stated in and the predictions made by a theory must be testable with empirical methods. Theories that are not testable, such as Freudian psychodynamics, cannot be classified as valid scientific theories. A theory also must be able to account for novel events within its domain. Finally, a good theory should be parsimonious. That is, it should explain a phenomenon with the fewest number of propositions possible.

Theories that are subjected to empirical tests can be confirmed or disconfirmed. Confirmation of a theory means that you have more confidence in the theory than before confirmation. Unfortunately, it is logically impossible to prove that a theory is absolutely correct. Theories that are disconfirmed may be modified or discarded entirely although many disconfirmed theories are adhered to for a variety of reasons.

In the course of testing a theory, various strategies can be used. Strong inference involves developing testable alternative explanations for a phenomenon and subjecting them simultaneously to an empirical test. The empirical test should be one that will unambiguously show which alternative is best. One way to test a theory is to use a confirmational strategy. That is, you design tests that will confirm the predictions made by the theory under test. When predictions are confirmed, then your confidence in the theory increases. Unfortunately, you may find confirming evidence even though the theory is wrong. Another approach is to adopt a disconfirmational strategy. In this case, you look for evidence that does not support the predictions made by a theory. Often the best strategy to adopt is to use both confirmational and disconfirmational strategies together.

Finally, a controversy exists over the role that a theory should play in driving research. Some scientists believe that research should be data driven, whereas others believe that research should be theory driven. Strong arguments have been made for each position, and no simple solution to the controversy exists.

## KEY TERMS

---

scientific theory

law

model

mechanistic explanation

functional explanation

quantitative theory

qualitative theory

descriptive theory

analogical theory

fundamental theory

domain

confirmational strategy

disconfirmational strategy

strong inference

## 3

## CHAPTER

## Getting Ideas for Research

## CHAPTER OUTLINE

**Sources of Research Ideas**

- Experience
- Theory
- Applied Issues

**Developing Good Research Questions**

- Asking Answerable Questions
- Asking Important Questions

**Developing Research Ideas: Reviewing the Literature**

- Reasons for Reviewing the Scientific Literature
- Sources of Research Information

**Performing Library Research**

- The Basic Strategy
- Using PsycINFO
- Using PsycARTICLES
- Other Computerized Databases
- General Internet Resources
- Computer Searching for Books and Other Library Materials
- Other Resources

**Reading a Research Report**

- Obtaining a Copy
- Reading the Literature Critically

**Factors Affecting the Quality of a Source of Research Information**

- Publication Practices
- Statistical Significance
- Consistency With Previous Knowledge
- Significance of the Contribution
- Editorial Policy
- Peer Review
- Values Reflected in Research

**Developing Hypotheses****Summary****Key Terms**

As a student who is just becoming acquainted with the research process, you are probably wondering just how to come up with good ideas for research. It may seem to you that, by this point in the history of psychology, every interesting research question must already have been asked and answered. Nothing could be further from the truth! Each year hundreds of novel research studies are published in scores of psychology journals. Or perhaps you do have some rather general idea of a topic that you'd like to explore but don't know how to convert that idea into something specific that you could actually carry out.

Once you learn how to go about it, finding a research topic and developing it into an executable project becomes relatively easy—you just have to know where and how to look. In fact, you may be surprised to find that your biggest problem is deciding which of several interesting research ideas you should pursue first. To help you reach that point, in the first part of this chapter we identify a number of sources of research ideas and offer some guidelines for developing good research questions.

Although finding and developing a research idea is usually the first step in the research process, the ultimate goal of that process, as noted in Chapter 1, is to develop valid explanations for behavior. These explanations may be limited in scope (e.g., an explanation of why a certain autistic child keeps banging his head against the wall) or comprehensive (e.g., a system that explains the fundamental mechanisms of learning). Of course, any single study will have only a limited purpose, such as to test a particular hypothesis, to identify how certain variables are related, or simply to describe what behaviors occur under given conditions. Yet each properly conceived and executed study contributes new information—perhaps, for example, by identifying new behaviors for which explanations will be needed or by ruling out certain alternative explanations. Ultimately, this information shapes the formulation of new explanations or tests the adequacy of existing ones.

In this chapter, we pursue two separate but related topics. First, we explore how to get research ideas and how to develop them into



viable, testable research questions. Second, we discuss how to do library research so that you can find research on the topic that interests you.

## SOURCES OF RESEARCH IDEAS

---

The sources of research ideas are virtually endless. They range from casual observation to systematic research. However, they can be seen as falling into three broad categories: experience, theory, and applied issues.

### Experience

Your everyday experience and observations of what goes on around you is a rich source of research ideas. Some of these observations may be unsystematic and informal. For example, after reading a newspaper article about a terrorist attack, you may begin to wonder how people who have to live with terrorism every day cope. Subsidiary questions might also come to your mind, such as: Do men and women cope differently with terrorism? Do adults adjust better than children? General questions like these can be translated into viable research questions. Other observations may be more systematic and formal. For example, after reading a journal article for a class, you may begin to formulate a set of questions raised by the article. These too could serve as the foundation of a viable research study.

**Unsystematic Observation** One of the most potent sources of research ideas is curiosity about the causes or determinants of commonplace, everyday behavior. You make a helpful suggestion to a friend, and she angrily rebukes you. Why? Perhaps she just found out she did not get the job that she wanted badly. Is this the cause, or is it something else? Or you study all week for an important exam, and the test results show you did very well. Although initially you feel good, the emotion soon passes, and you find yourself falling into a deep depression. What caused this seemingly strange shift in your emotions? Such observations can provide the basis for a research project.

Casual observation of animal behavior also can lead to research ideas. Behaviors such as starlings staging a mass attack on a soaring hawk, a squirrel dropping an acorn on your head, and the antics of a pet all raise questions about why those behaviors occur—questions that can be the basis of a research idea. For example, Niko Tinbergen's (1951) well-known research on territorial defense and courtship behavior in the three-spined stickleback (a minnow-sized fish that inhabits European streams) began when Tinbergen happened to observe some odd behavior in a small group of sticklebacks that he kept in an aquarium near a window. During breeding season the males' underbellies turn bright red and the males construct a nest on the bottom of the stream. They then defend the territory around the nest from intrusion by other male sticklebacks. One day as a Dutch mail truck passed by the window, Tinbergen watched in astonishment as the male sticklebacks rushed to the surface of the water nearest the window in an apparent attempt to attack the red truck and drive it away. Because mail trucks are not normally a part of a stickleback's environment, Tinbergen wondered whether the males' red underbellies might be the normal trigger for attack

by other males. This was the catalyst for a carefully designed research project aimed at answering this question. (See Chapter 4 for more information about this research.)

Unsystematic observation sometimes is a good way to discover a general research idea. Given your casual observations, you may decide to study a particular issue. For example, your questions about coping with terrorism may lead you to some general questions about the factors that cause terrorism. You may decide to focus your research on one or two variables that you believe are strongly associated with those decisions. For example, you could focus your research on the attitudes that underlie terrorism and how religion and terrorism relate.

You also can get research ideas just by paying attention in your classes. In many classes, your professors undoubtedly use research examples to illustrate points. As you listen to or read about these research examples, you may be able to think of some interesting research questions. For example, you might ask whether the research results just presented apply equally to men and women or to Western as well as non-Western cultures. With a little follow-up digging through published research, you may find that many questions surrounding gender and culture remain wide open.

Here is a good example of how this works. In my (Bordens) social psychology class, students read an article by H. Andrew Sagar and Janet Schofield originally published in the *Journal of Personality and Social Psychology* (1980). The article reports an experiment conducted by Sagar and Schofield on how behavior of Black and White children is perceived. In their experiment, 40 Black and 40 White children were shown an artist's rendering of four different situations depicting two children (e.g., one child poking another in a classroom). Each picture was accompanied by an oral description. The oral description for the "poking" picture was as follows:

Mark was sitting at his desk, working on his social studies assignment, when David started poking him in the back with the eraser end of his pencil. Mark just kept on working, David kept poking him for a while, and then he finally stopped. (Sagar & Schofield, 1980, p. 593)

The researchers manipulated the race of the child engaging in the behavior (Black or White) and the race of the victim (Black or White). For example, in one version David (the "actor") was Black and Mark (the victim) was White. In another version, Mark was Black and David was White. Participants rated the degree to which several adjectives describing the actor's behavior applied to the situation (e.g., playful, mean, friendly, or threatening). The results showed that participants rated the actor's behavior as more threatening and mean when the actor was Black than when the actor was White. So the same behavior was rated differently depending upon the race of the actor.

By itself, this finding is interesting. However, just as interesting is the number of questions this study raises that could serve as the foundation for further experiments. In discussions of this article, students invariably bring up a number of issues that could be studied empirically. For example, students often ask if the results are the same for male and female children. Since Sagar and Schofield (1980) did not include participant gender as a variable, we have no way of knowing. It is an open question. Another question is whether the results would be the same if the actor belonged to another ethnic or racial group (e.g., Asian or Hispanic). Once again, Sagar and

Schofield did not evaluate this, so we don't know. Finally, students note that the study was published in 1980. They wonder if the results are still valid today. Unfortunately, nobody has ever replicated Sagar and Schofield's study. So, once again, we just don't know. You could use these questions, and a myriad of others, to develop research ideas for a number of studies.

Casual observations are only a starting point. You still must transform your casual observations into a form that you can test empirically. Rarely will you be able to infer the causes of observed behavior from your casual observations. You can only infer such causes through a careful and systematic study of the behavior of interest.

**Systematic Observation** Systematic observation of behavior is another powerful source of research ideas. In contrast to casual observation, systematic observation is planned. You decide what you are going to observe, how you are going to observe it, and how you will record your observations. Your own systematic observations of real-world behavior can provide the basis for a research idea. Consider the work of Jean Piaget (1952). Piaget spent many an hour systematically observing the behavior of his own children at home and other children on playgrounds. These observations helped lay the foundation for his comprehensive theory of cognitive development.

It is important to note that Piaget did not make his observations in a vacuum. Instead, he approached a situation with some ideas in mind about the nature of children's thought processes. As he observed children's behavior, he began developing hypotheses that he tested in further research.

A second valuable source of systematic observation is published research reports. Instead of observing behavior firsthand, you read about other firsthand observations from researchers. Published research offers an almost limitless source of systematic observations of both human and animal behavior made under well-defined conditions. Although such research answers many questions, it typically raises more than it answers. Are the results reliable? Would the same thing happen if participants with different characteristics were used? What is the shape of the function relating the variables under study? Would you obtain the same results if the dependent measure were defined differently? These questions and others like them provide a rich source of research ideas.

Another potent, systematic source of research ideas is your own previous or ongoing research. Unexpected observations made during the course of a project (e.g., a result that contradicts expectations) or the need to test the generality of a finding can be the basis for further research. As you examine your data, you may see unexpected relationships or trends emerging. These trends may be interesting enough to warrant a new study.

For example, my (Bordens) research colleague and I conducted an experiment on the effect of the number of plaintiffs in a civil trial on damage awards. In our original experiment (Horowitz & Bordens, 1988), we found that as the size of the plaintiff population increased so did damage awards. This finding then led us to wonder what number of plaintiffs yields the highest award. In follow-up experiments we found that the critical number of plaintiffs was four.

In this example, we found something interesting (increasing the size of the plaintiff population leads to higher damage awards), which led to another interesting

question (what is the critical number?). In the same way, you can get research ideas from your own research.

It is important to note that this particular source of research ideas usually is not immediately available to the scientific community. Other researchers may not become aware of your findings until you publish or present them. Consequently, you and your close colleagues may be the only ones who can benefit from this potentially rich source of research ideas.

Finally, you may be able to get some research ideas by perusing research projects being run on the Internet. At any given time there may be many different psychological research projects being conducted there. These include nonexperimental studies such as surveys, as well as experimental studies. You can find a wide variety of such studies on the Hanover College Psychology Department's *Psychological Research on the Net* Web site (at the time of this writing, <http://psych.hanover.edu/research/exponnet.html>). This Web site lists psychological studies broken down into categories (e.g., social psychology, cognition, and personality). You can take part in these studies, and you may get some good ideas for your own research based on your participation.

## QUESTIONS TO PONDER

---

1. How can experience help you come up with research ideas?
2. How can unsystematic observation help you develop research ideas?
3. How can systematic observation help you develop research ideas?

## Theory

As defined in Chapter 2, a *theory* is a set of assumptions about the causes of behavior and rules that specify how those causes act. Designed to account for known relationships among given variables and behavior, theories can also be a rich source of research ideas.

Theories can lead to the development of research questions in two ways. First, a theory allows you to predict the behavior expected under new combinations of variables. For example, *terror management theory* (Solomon, Greenberg, & Pyszczynski, 1991) suggests that when you become aware that you live in an unpredictable world in which your existence could end at any moment, you get scared and experience “terror.” The theory also predicts that you develop a variety of strategies to cope with your mortality as a way of managing the terror. The theory predicts that cultures provide “terror shields” that buffer us against sources of terror, most notably our own mortality. One such terror shield is to begin thinking about positive things to counter the negative emotions associated with mortality (DeWall & Baumeister, 2007). In fact, DeWall and Baumeister conducted a series of experiments looking at how positive emotions reduce anxiety generated by facing one's mortality. Let's see how this all worked.

DeWall and Baumeister (2007) hypothesized that after facing the prospect of death people begin an unconscious search for positive, emotionally pleasant

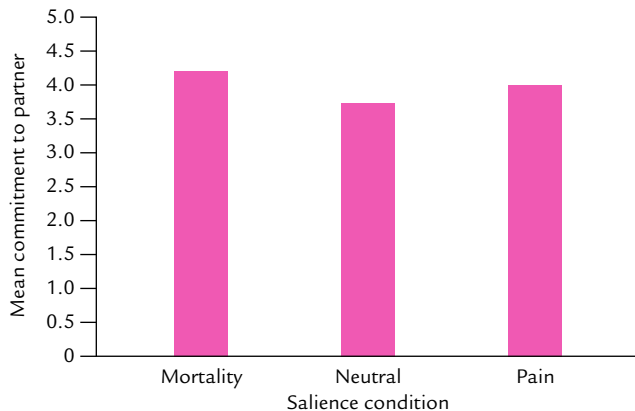
information. According to DeWall and Baumeister, “clutching at happy thoughts may serve the function . . . of preventing the conscious mind from being paralyzed by the terror of death” (p. 984).

Before we examine DeWall and Baumeister’s study and results, let’s pause and review how their research idea flowed from a theory. They started with three postulates from terror management theory: Each of us is mortal, individuals are frightened (terrorized) by knowledge of their own mortality, and those individuals will find ways of managing terror. They reasoned that one way to counter the terror is to “think happy thoughts.” So, based on terror management theory, they developed the research hypothesis discussed earlier. The hypothesis flowed directly from the predictions of terror management theory. Now back to the study . . .

In their first experiment, 64 males and 141 females participated. Participants completed several measures, including items concerning their own mortality. DeWall and Baumeister manipulated the wording of the questions to create two experimental conditions. In the “mortality salience” condition the questions evoked thoughts of the participants’ own death. Participants in the “mortality neutral” condition answered questions evoking unpleasant thoughts that were unrelated to death. Next, participants completed a word completion task which included several words that could be completed in a positive or neutral way (e.g., jo\_ could be completed as either joy or jog) or in a negative or neutral way (e.g., ang\_\_ could be completed as either anger or angle).

The results were consistent with the predictions from terror management theory. As shown in Figure 3-1, participants in the “mortality salience” completed more words in a positive direction than the participants in the “mortality neutral” condition. This finding supports the predictions derived from terror management theory.

The second way that theory can generate research ideas arises when two or more alternative theories account for the same initial observations. This situation may provide a fascinating opportunity to pit the different interpretations against one another. If the alternatives are rigorously specified and mutually exclusive, they may lead to



**FIGURE 3-1** The relationship between the salience condition and mean commitment to one’s romantic partner.

SOURCE: DeWall and Baumeister, 2007.

different predictions about what will be observed under a new set of conditions. In this case, a single experiment or observation may be enough to provide strong support for one alternative over another.

One example of this source for research ideas is the different accounts for attitude change provided by cognitive dissonance theory (Festinger, 1957) and self-perception theory (Bem, 1972). *Cognitive dissonance theory* maintains that when there is inconsistency between our attitudes and our behavior, a negative motivational state called *cognitive dissonance* arises. Because this is a negative state, dissonance theory states that an individual will be motivated to reduce or eliminate it through attitude or behavior change. The linchpin of dissonance theory is the arousal of cognitive dissonance. It is a necessary precondition for attitude change. Without dissonance, no attitude change should occur.

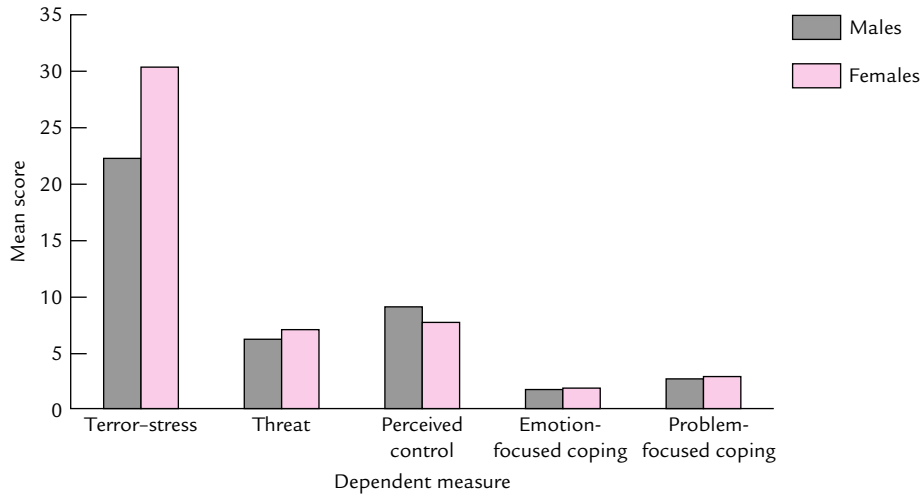
In contrast, *self-perception theory* states that dissonance is not necessary for attitude change. Instead, the theory states that we learn about our motives by observing and evaluating our own behavior. In short, the theory maintains that we observe our own behavior and then assume that our attitudes must be consistent with that behavior. So, if we behave in a manner that is inconsistent with an attitude, we change the attitude so that it is consistent with our self-observed behavior. Attitude change comes about because we *reason* that we have a particular attitude that is consistent with our behavior and not because we are motivated to reduce cognitive dissonance.

Here we have an example of two theories designed to account for the same behavior. Which one is correct? This is where research comes in. The question of when, or if, either or both theories account for behavior is an empirical one. When researchers addressed this question, they found that both theories were valid. There are situations in which dissonance is clearly aroused, and it motivates attitude change. There are other situations in which we undergo attitude change without dissonance arousal. Situations like this provide a fruitful source of research ideas.

### Applied Issues

Often research ideas arise from the need to solve practical problems. Chapter 1 distinguished between basic and applied research. Applied research is problem oriented whereas basic research is aimed toward building basic knowledge about phenomena. You might design an applied research study to develop interventions to help people cope with terrorism. Of course, before you can design any intervention, you must first know something about how people react to terrorism. It may be that you have to use different intervention programs depending on individuals' unique characteristics.

A study by Moshe Zeidner (2006) investigated how Israeli men and women reacted to the chronic threat of terrorism. Male and female Israelis living in and around Haifa (in northern Israel) completed several measures designed to determine how they coped with the chronic threat of terrorism. Participants completed a battery of measures that included four categories of variables: (1) terror stress (measuring reactions to continued political violence and conflict), (2) personal variables (experience of negative emotion and degree of control over events), (3) coping processes (strategies used to cope with terrorism), and (4) stress reactions (symptoms experienced). Zeidner collected his data between April and June 2002, which was at the height of the Palestinian



**FIGURE 3-2** Gender differences in coping with chronic terrorism.  
SOURCE: Based on data from Zeidner, 2006.

al-Aqsa Intifada. Zeidner's results were that men and women had different responses to the threat of terrorism. As shown in Figure 3-2, Israeli women experienced much higher levels of terror-related stress than men experienced and reported feeling more threatened by terrorism than men. Further, men indicated slightly more perceived control over the situation than did women. Women reported using more emotion-focused (e.g., denial, behavioral and mental disengagement, and alcohol/drug use) and problem-focused (e.g., positive reinterpretation and social support) behavior than men.

A wealth of other practical problems lend themselves to similar research solutions. For example, finding an effective way to get people to practice safe sex or finding an effective diet that people will follow might require a systematic evaluation of several proposed solutions to identify those that lead to success. Applied research also might identify the most effective therapy for depression or develop a work environment that leads to the highest levels of productivity and job satisfaction. Thus, a need to solve a practical problem can be a rich source of research ideas.

## QUESTIONS TO PONDER

---

1. In what two ways can a theory help you develop research ideas?
2. How can applied issues suggest research ideas to you?

## DEVELOPING GOOD RESEARCH QUESTIONS

---

Coming up with a creative and unique general research question based on experience, theory, or application is not sufficient in science. After coming up with an inspired idea, you must translate that idea into a good research question that can be

tested empirically. This section describes how to identify good research questions and suggests what kinds of questions are likely to be important.

### Asking Answerable Questions

The first step in developing a workable research project is to ask the kind of question that you can answer with the scientific method. Not all questions can. Here are a few questions that cannot be answered by scientific means: Does God exist? Why is there suffering in the world? Are there human abilities that cannot be measured? How many angels can stand on the head of a pin? Is embryonic stem cell research moral or immoral?

**Asking Empirical Questions** The preceding questions are not answerable by scientific means because you can't find the answers through objective observation. To be objective a question must meet three criteria. First, you must be able to make the observations under precisely defined conditions. Second, your observations must be reproducible when those same conditions are present again. Third, your observations must be confirmable by others.

A question you can answer with objective observation is called an **empirical question**. Here are some examples of empirical questions: Do males and females cope differently with terrorism? Do men and women prefer different characteristics in potential mates? Does a deprived early environment result in lower intelligence? Is punishment an effective tool in socializing children? You can answer all of these questions through appropriately designed and executed research. Unlike the first set of questions, the second set identifies variables that you can define in terms of observable characteristics. For example, the question of whether males and females cope differently with terrorism asks about the relationship between two observable variables: gender and coping skills.

Some questions seem to be empirical but are formulated too broadly to make appropriate observations. Consider the following example of such a question: Do children raised in a permissive atmosphere lack self-discipline as adults? Before you can answer this question, a number of preliminary questions must be addressed. What exactly is a permissive atmosphere? How do you measure permissiveness? Precisely what does it mean to lack self-discipline, and how do we determine when self-discipline is present or absent? Until you can specify exactly what these terms mean and how to measure the variables they represent, you cannot answer the original question.

**Operationally Defining Variables** One way to give precise meaning to the terms that you use is to provide an operational definition for each variable you are using. An **operational definition** involves defining a variable in terms of the operations required to measure it. Defining variables operationally allows you to measure precisely the variables that you include in your study and to determine whether a relationship exists between them. For example, you could operationally define "permissive parenting" in terms of the frequency that parents discipline their children for bad behavior. You could operationally define "lack of self-discipline" in an adult as the number of reprimands a person receives at work for late or sloppy work. With these precise



definitions, you are now in a position to conduct a study to see if increased parental permissiveness is related to decreased adult self-discipline in the hypothesized way.

Although defining variables operationally is generally a good thing, there is a downside. Operational definitions restrict the generality of answers obtained. Permissive parenting, for example, is no longer addressed in general but as particular behaviors defined as permissive. Self-discipline is no longer addressed in general but in the context of specific behaviors said to indicate self-discipline. Other ways of measuring the two variables may yield a different answer to the question. Nevertheless, without using operational definitions, the question cannot be answered meaningfully.

To summarize, to conduct meaningful research you must choose a question that you can answer through scientific means. You must then operationally define your variables carefully so that you are working with precise definitions. When you have formulated your empirically testable question, you then proceed to the next step in the research process.

### Asking Important Questions

Developing answerable questions is not enough. They also should be important questions. Researching a question imposes demands on your time, financial resources, and the institution's available space. Researching a question makes demands on the available population of human participants or animal subjects. You should not expend these resources to answer trivial questions. However, whether a question is important is often difficult to determine. It is certainly possible to obtain valuable information in the course of answering an apparently unimportant question. Some questions that once seemed important to answer now appear trivial. Some rough guidelines will help in identifying important questions.

A question is probably important if answering it will clarify relationships among variables known to affect the behavior under study. For example, knowing that memory tends to deteriorate with time since learning, you would want to establish the rate of deterioration as a function of time. You would want to identify how the amount of initial practice, overlearning, previous learning, activity during the retention interval, and other such factors combine to determine the rate of forgetting under specified conditions.

A question is probably important if the answer can support only one of several competing models or theoretical views. As noted in Chapter 2, developing and testing such questions is at the heart of the scientific method. The answers to such questions allow you to "home in" on a proper interpretation of the data. (We discuss this technique later in the chapter.) On the negative side, if the theories under test are later discarded, research designed to test the theories may become irrelevant unless the findings demonstrate clear empirical relationships that other theories must explain.

A question is probably important if its answer leads to obvious practical application. (However, the lack of obvious practical application does *not* render the question automatically unimportant!) Researchers have conducted much research to identify working conditions that maximize productivity and job satisfaction or to screen drugs for potential effectiveness in controlling psychosis. Few would argue that the answers to these problems are unimportant.

In contrast, a question is probably unimportant if its answer is already firmly established. *Firmly established* means that different scientists have replicated (duplicated) a research finding and agree that the finding does occur under the stated conditions. Unless you can identify serious deficiencies in the methods used to establish those answers, performing the research again is likely to be a waste of time.

A question is probably unimportant if the variables under scrutiny are known to have small effects on the behavior of interest and if these effects are of no theoretical interest. A question is also probably unimportant if there is no a priori reason to believe that the variables in question are causally related. Research aimed at determining whether the temperature of a room affects memory recall for faces may turn out to have surprising and useful results. However, without a reason to expect a relationship, such research amounts to a “fishing expedition” that is unlikely to pay off. Your time would be better spent pursuing more promising leads.

When you have identified your research idea, the next step is to develop it to the point at which you can specify testable hypotheses and define the specific methods to be used to test these hypotheses. You accomplish this step by identifying and familiarizing yourself with research already conducted in your area of interest, an activity called “reviewing the literature.” We show you how to review the literature and evaluate research reports in the following section.

## QUESTIONS TO PONDER

---

1. What are the characteristics of an empirical question?
2. Why is it necessary to define your terms operationally?
3. What makes a research question important and why should you ask important questions?

## DEVELOPING RESEARCH IDEAS: REVIEWING THE LITERATURE

---

One of the most important preliminary steps in the research process is doing a thorough review of the scientific literature on the topic that you have identified for study. This is true whether you begin only with a vague idea of a research project or with a well-developed research plan. In this section, we discuss the tools, techniques, and knowledge that will enable you to identify, read, and evaluate published information on your research topic. In addition, we discuss the process of scientific peer review and describe how this process affects the content and quality of published scientific findings.

### Reasons for Reviewing the Scientific Literature

A **literature review** is the process of locating, obtaining, reading, and evaluating the research literature in your area of interest. Perhaps the most important reason for

conducting a literature review is to avoid needless duplication of effort. No matter what topic you choose, chances are that someone has already done research on it. By becoming familiar with that area through a literature review, you can avoid “reinventing the wheel.”

Another reason for conducting a literature review is that your specific research question may have already been addressed and answered. If so, then conducting your research as originally planned would be a waste of time. This does not mean, however, that you must start over from scratch. To the contrary, your literature review may reveal other questions (perhaps more interesting) that remain to be answered. By familiarizing yourself with existing research and theory in an area, you can revise your research project to explore some of these newly identified questions.

Another reason for reviewing the literature applies to the design phase of your research. Designing a study involves several decisions as to what variables to include and how to measure them, what materials or apparatus to use, what procedures to use, and so on. Published research provides you with a rich resource for addressing these important design questions. You may find, for example, that you can use established procedures and existing materials.

Reviewing the literature also keeps you up to date on current empirical or theoretical controversies in a particular research area. As science progresses, new ideas develop concerning age-old behavioral issues. For example, there is a debate concerning the motives for altruistic behavior. Some argue that empathy (a concern for the victim) motivates altruism and others argue that egoism (self-satisfaction) motivates altruism. Such controversies not only provide a rich source of research ideas but also give direction to specific research hypotheses and designs.

### Sources of Research Information

Sources of information about a topic range in quality from the high levels found in the scholarly books and journals of a discipline to the low levels found in the supermarket tabloids of the sensationalist press. Although information presented in the tabloids may arouse your curiosity and suggest a topic for scientific research, you cannot count on that information to be accurate or even true. Popular writing found in magazines such as *Newsweek* may provide more reliable information gleaned from scientific sources, but the information presented generally lacks the detail that would allow you to determine much beyond the major conclusions offered. More substantive writing aimed at a better-educated reader generally provides more details about the methods used to gather the information but still omits important details and may not mention alternative interpretations or other evidence for or against the conclusions presented. You can only count on scholarly sources to provide the level of detail and thoroughness needed for a competent scientific review. Table 3-1, which is based on an analysis provided by the Cornell University Library (2000), identifies four types of periodicals and compares them on a number of important features. You can use this table to help you determine whether a publication is scholarly or not.

**TABLE 3-1 Comparison of Four Types of Published Periodicals**

SCHOLARLY	SUBSTANTIVE NEWS/ GENERAL INTEREST	POPULAR	SENSATIONAL
Sober, serious look with graphs and tables	Attractive appearance, usually with photographs	Often have a slick, attractive appearance with many photographs	Often in newspaper format
Reference citations always provided	Sources are sometimes cited	Sources are rarely, if ever, cited	References to sources are often obscure
Written by a scholar in the field or someone who has done research in the field	Articles written by members of editorial staff, scholar, or freelance writer	Written by a wide range of authors who may or may not have expertise in an area	Written by a variety of authors
Language of the discipline, assuming a scholarly background of the reader	Language geared to educated audience, but no specialty assumed	Written in simple language with short articles geared to audience with minimal education	Elementary, inflammatory language geared to a gullible audience
Report original research	Do not report original research, report <i>on</i> research in format geared to a general audience	Research may be mentioned, but it may come from an obscure source	Support may come from pseudoscientific sources
Many, but not all, published by professional organizations	Published by commercial publishers or individuals, but some from professional organizations	Published commercially with the intent to entertain the reader, sell products, or promote a viewpoint	Commercially published to arouse curiosity and play to popular superstition. Use flashy, astonishing headlines
Examples: <i>Journal of Personality and Social Psychology, Child Development, Journal of Experimental Psychology</i>	Examples: <i>National Geographic, Scientific American, New York Times, Christian Science Monitor</i>	Examples: <i>Time, U.S. News &amp; World Report, Newsweek, Parents, Reader's Digest</i>	Examples: <i>National Enquirer, Globe, Star, Weekly World News</i>

SOURCE OF MUCH OF THE INFORMATION: Cornell University Library Web site.

## QUESTIONS TO PONDER

---

1. Why should you conduct a literature review before you begin to design your study?
2. What are the differences between the different types of periodicals, and on which should you rely most heavily (and why)?

Sources of research findings include books, scholarly journals, conventions and professional meetings, and others such as personal communications and certain pages on the World Wide Web. Here are a few things you should know about these sources.

**Primary Versus Secondary Sources** Sources containing research information are classified according to whether a source is primary or secondary. A **primary source** is one containing the full research report, including all details necessary to duplicate the study. A primary source includes descriptions of the rationale of the study, its participants or subjects, materials or apparatus, procedure, results, and references. A **secondary source** is one that summarizes information from primary sources (such as presenting the basic findings). Secondary sources of research include review papers and theoretical articles that briefly describe studies and results, as well as descriptions of research found in textbooks, popular magazines, newspaper articles, television programs, films, or lectures. Another type of secondary source is a meta-analysis. In a meta-analysis, a researcher statistically combines or compares the results from research in a particular area to determine which variables are important contributors to behavior. (We discuss meta-analysis in Chapter 8.)

Distinguishing primary from secondary sources is important. Students often rely too heavily on secondary sources, perhaps because it can be a daunting task to read a primary source research report. The language can be technical, and the statistical tests reported can be intimidating. However, with some experience and perseverance, you can get through and understand primary source materials.

Another reason that students may rely heavily on secondary sources is to “save time.” After all, someone else has already read and summarized the research, so why not save time and use the summary? This sounds good but can lead to trouble. The author of a secondary source may describe or interpret research results incorrectly or simply view data from a single (and perhaps narrow) theoretical perspective. In addition, secondary sources do not usually present detailed descriptions of methods used in the cited studies. You must know the details of the methods used in a study before you can evaluate its quality and importance. The only way to obtain such detailed information is to read the primary source.

Relying heavily on secondary sources can be dangerous. You cannot be sure that the information in a secondary source is complete and accurate. In a study of this issue, Treadway and McCloskey (1987) found that many secondary sources had misrepresented the methods and results of a classic experiment conducted by Allport and Postman (1945). These representations led researchers and sometimes courts to draw incorrect inferences concerning the role of racial bias in eyewitness accuracy. To avoid this trap, obtain and read the original report.

Secondary sources do have value, which lies in the summaries, presentations, and integrations of results from related research studies. The secondary source provides an excellent starting point for your literature search. Additionally, an up-to-date review paper or meta-analysis includes a reference section from which you can generate a list of primary sources. However, you should not consider a secondary source as a substitute for the primary source.

You may need to use a secondary source if the primary source it refers to is not available. If you must do so, always stay aware of the possible problems. In addition, cite the secondary source in your research report, not the primary one that you were unable to obtain.

To summarize, use secondary sources as a starting point in your literature search. Avoid overreliance on secondary sources and make every effort to obtain the primary sources of interest to you. Only by reading the primary source can you critically evaluate a study and determine whether the reported results are reliable and important. Finally, do not rely on a single secondary source. The author of a review article may not have completely reviewed the literature. Always augment the information obtained from a secondary source with a thorough literature search of your own.

**Books** You are probably most familiar with general textbooks (such as those covering introductory psychology) or texts covering content areas (such as motivation and emotion, abnormal psychology, social psychology, personality, or cognitive psychology). Other books may contain more detailed information about your research topic. Specialized professional texts present the results of programmatic research conducted by the author over a period of years. These specialized texts may cover research previously published in journals, as well as findings not presented elsewhere. Edited anthologies present a series of articles on related topics, each written by a different set of authors. Some anthologies are collections of articles previously published separately; others present articles written especially for the occasion. Either kind of text may present reviews of the literature, theoretical articles, articles dealing with methodological issues, or original research.

Anthologies are useful because they assemble papers that the editor believes are important in a given area. However, be cautious when reading an anthology. The editor may be biased in judgment on which articles to include. Also, be sure to check the original publication date of articles in an anthology. Even if the publication date of the anthology is recent, it may contain outdated (sometimes classic) articles.

Texts or anthologies are most valuable in the early stages of the literature search. Often you can use the references from these books to track down relevant articles. You may have to treat books (especially textbooks) as secondary sources. Whenever you use a textbook as a source, make an effort to obtain a copy of the primary source cited in the textbook.

The articles in an anthology may be original works and thus can be treated as primary sources—provided that they have been reproduced exactly, not edited for the anthology. Be careful about relying on a chapter reproduced from a book. Isolating a single chapter from the original book can be misleading. In other chapters from the same book, the original author might elaborate on points made in the reproduced chapter. You could miss important points if you do not read the original work.

Whereas some books present original research, others provide only summaries. For example, if you were studying the development of intelligence, you could use Piaget's *The Origins of Intelligence in Children* (1952) as a good original source. However, a book such as *Piaget's Theory of Cognitive Development* by Wadsworth (1971)—a primer on Piaget's theory—should be treated as a secondary source in which you may find references for Piaget's original work.

Whatever route you choose, keep in mind one important factor. Even though you may have used an original work such as Piaget's (1952), problems with using it as a principal source may still exist. Books (especially by noted authors) may not undergo as rigorous a review as works published in scientific journals. You cannot be assured of the quality of any original research reported in the book. In addition, you would be well advised to seek out recent research on the issues covered in a book. Was Piaget correct when he speculated in his book about the origins of intelligence? Research published since his book came out may bear on this question. A review of the recent research would help you evaluate Piaget's theory and contributions.

## QUESTIONS TO PONDER

---

1. What is the difference between a primary and a secondary source, and why should you not rely too heavily on secondary sources?
2. What are the advantages and disadvantages to using various types of books as sources?

**Scholarly Journals** Although textbooks are valuable, the information they contain tends to be somewhat dated. By the time a scientific finding makes its way into a text, it could already have been around for several years. For current research and theories regarding a subject, researchers turn to scholarly journals. Like popular magazines, journals appear periodically over the year in monthly, bimonthly, or quarterly issues. Some journals focus on detailed research reports (although occasionally a theoretical or methodological article may appear). These research reports are the most important primary sources. Other journals deal with reviews of the literature, issues in methodology, or theoretical views.

Table 3-2 provides a list of journals currently published by the American Psychological Association (APA), the Society for Psychological Science, and the Psychonomic Society. (The list is not complete. In addition to those listed, many journals are published by major textbook publishers. You become familiar with these by doing reviews of the literature.)

Keep in mind that not all journals are created equal. You must consider the source. When you submit your work to a **refereed journal**, it is reviewed, usually by two (or more) reviewers. Other, **nonrefereed journals** do not have such a review procedure; the articles may be published in the order in which they were received or according to some fee that the author must pay. The review process is intended to ensure that high-quality articles appear in the journal. Although problems do occur

**TABLE 3-2 Journals Published by Major Psychological Organizations****JOURNALS OF THE AMERICAN PSYCHOLOGICAL ASSOCIATION**

*American Journal of Orthopsychiatry*  
*American Psychologist*  
*Behavioral Neuroscience*  
*Consulting Psychology Journal: Practice and Research*  
*Cultural Diversity and Ethnic Minority Psychology*  
*Developmental Psychology*  
*Emotion*  
*Experimental & Clinical Psychopharmacology*  
*Group Dynamics: Theory, Research, and Practice*  
*Health Psychology*  
*History of Psychology*  
*Journal of Abnormal Psychology*  
*Journal of Applied Psychology*  
*Journal of Comparative Psychology*  
*Journal of Consulting and Clinical Psychology*  
*Journal of Counseling Psychology*  
*Journal of Educational Psychology*  
*Journal of Experimental Psychology: Animal Behavior Processes*  
*Journal of Experimental Psychology: Applied*  
*Journal of Experimental Psychology: General*  
*Journal of Experimental Psychology: Human Perception and Performance*  
*Journal of Experimental Psychology: Learning, Memory, and Cognition*  
*Journal of Family Psychology*  
*Journal of Occupational Health Psychology*  
*Journal of Personality and Social Psychology*  
*Journal of Psychotherapy Integration*  
*Neuropsychology*  
*Prevention & Treatment*  
*Professional Psychology: Research and Practice*  
*Psychoanalytic Psychology: A Journal of Theory, Practice, Research, and Criticism*  
*Psychological Assessment*  
*Psychological Bulletin*  
*Psychological Methods*  
*Psychological Review*  
*Psychology and Aging*  
*Psychology of Addictive Behaviors*  
*Psychology of Men and Masculinity*  
*Psychology, Public Policy, and Law*  
*Psychotherapy: Theory, Research, Practice, Training*  
*Rehabilitation Psychology*  
*Review of General Psychology*



**TABLE 3-2 Journals Published by Major Psychological Organizations** *continued*

JOURNALS OF THE SOCIETY FOR PSYCHOLOGICAL SCIENCE
<i>Current Directions in Psychological Science</i> <i>Psychological Science</i> <i>Psychological Science in the Public Interest</i>
JOURNALS OF THE PSYCHONOMIC SOCIETY
<i>Behavior Research Methods, Instruments, &amp; Computers</i> <i>Cognitive, Affective, &amp; Behavioral Neuroscience</i> <i>Learning &amp; Behavior</i> (formerly <i>Animal Learning &amp; Behavior</i> ) <i>Memory &amp; Cognition</i> <i>Perception &amp; Psychophysics</i> <i>Psychonomic Bulletin &amp; Review</i>

with the review procedures, you can have greater confidence in an article from a refereed journal than in one from a nonrefereed journal.

A problem you are more likely to encounter in a nonrefereed journal than in a refereed journal is information that is sketchy and incomplete (Mayo & LaFrance, 1977). If information is incomplete, you may not be able to determine the significance of the article. Rely more heavily on articles published in high-quality, refereed journals than on articles in lower-quality, nonrefereed journals.

How do you know if a journal is refereed or nonrefereed? Check the inside front or rear cover of an issue of most journals for a statement of the journal's review policy. For example, on the inside rear cover of the journal *Psychological Science*, under "Information for Contributors," you can find the journal's review policy. It states that manuscripts are reviewed by two members of the editorial team. Thus, *Psychological Science* is a refereed journal.

You can assess the quality of a research journal in several ways. First, you can consult *Journals in Psychology*, published by the APA. This publication lists journals alphabetically and gives their manuscript acceptance rates. Top journals in a field have low acceptance rates (15% or less), whereas lesser journals have higher acceptance rates. Second, you can consult the *Journal Citations Report* available online from the Institute for Scientific Information (ISI) *Web of Knowledge*. Journals are ranked within category by their *impact factor*, which is a measure of "the frequency with which the 'average article' in a journal [was] cited in a particular year . . ." (Institute for Scientific Information [ISI], 1988, p. 10A). Third, you can consult the *Social Science Citations Index (SSCI)*. One section of this publication lists journals by category (e.g., psychology) and subcategory (social psychology, experimental psychology, etc.). Fourth, you can use the method of authority discussed in Chapter 1. Ask your professors which journals in their fields of specialty are of highest and lowest quality.

## QUESTIONS TO PONDER

---

1. Why are scholarly journals the preferred sources for research information?
2. What is the difference between a nonrefereed and a refereed journal? Which is more trustworthy (and why)?
3. How do you assess the quality of a scholarly journal?

**Conventions and Professional Meetings** Books and journals are not the only sources of research findings, nor are they necessarily the most current. Behavioral scientists who want the most up-to-date information about research in their areas attend psychological conventions. If you attended one of these conventions, you would find a number of **paper sessions** covering different areas of research. Paper sessions are usually simultaneously conducted in different rooms and follow one another throughout the day (much as classes do on campus).

When you register at a convention, you receive a program listing the times and places for each session. Figure 3-3 shows a page from the program of the 2009 meeting of the Midwestern Psychological Association. Listed under the session shown are the times at which the papers will be presented, the titles of the papers, the names of the authors, and short abstracts of the papers. You can use the program to identify papers relevant to your research interests. Each participant at a paper session is allotted time to describe his or her most recent findings and then usually has about 5 minutes to answer any questions from the audience.

Paper sessions are not the best way to convey details of methodology. The written report is far superior for that purpose. At a convention, the author of a paper typically has only 15 minutes to describe his or her research. In that short time, the author must often omit some details of methodology.

An increasingly popular format for convention presentations is the **poster session**. In this format, the presenter prepares a poster that is displayed on a bulletin board. The poster includes an introduction to the topic and method, results, discussion, and reference sections, and the presenter is usually there to discuss the research with you and answer any questions. This forum allows the author to provide more details than would be practical in a paper session and allows you to speak directly to the researcher about the research. Many good research ideas can emerge from such encounters.

Attending a paper or poster session has two distinct advantages over reading a journal article. First, the information is from the very frontiers of research. The findings presented may not appear in print for many months (or even years), if ever. Attending a paper session exposes you to newly conducted research that might otherwise be unavailable to you. Second, it provides an opportunity to meet other researchers in your field and to discuss ideas, clarify methodology, or seek assistance. These contacts could prove valuable in the future.

One drawback to paper and poster sessions at a convention is that a convention can be expensive to attend. In most instances, conventions are located in cities other than where you live. This means you must pay for travel, lodging, and food. Fortunately, you can gain some of the benefits of going to a conference by obtaining a copy

**ATTITUDES & PERSUASION**  
\*\*\*\*\*

*Thursday, 10:00–12:00* *Salon 5 & 8*  
DUANE WEGENER, Purdue University, Moderator

**10:00**  
**Going With Your Gut: Attitudes and BMI Predict Eating Enjoyment**  
ALLEN R. MCCONNELL, Miami University; SARA N. AUSTIN, Miami University;  
ELIZABETH W. DUNN, University of British Columbia; CATHERINE D. RAWN, University of  
British Columbia  
mcconnar@muohio.edu  
We explored how one's body (operationalized as body mass index), in addition to one's implicit  
and explicit attitudes, predicts one's enjoyment of eating chocolates and apples. Although all  
three indexes predicted enjoyment, BMI made a unique contribution above one's attitude mea-  
sures, suggesting a role for embodied knowledge in predicting behavior.

**10:15**  
**Implicit Theories of Judgment: Effects on Attitudes and Evaluation**  
CLIFFORD D. EVANS, Miami University; AMANDA B. DIEKMAN, Miami University  
evanscd1@muohio.edu  
This study examined the effect of naïve theories about judgment on attitudes and evaluative out-  
comes. Explicit information influenced implicit and explicit attitudes for feelings-based theorists,  
but influenced only explicit attitudes for reasons-based theorists. Implicit attitudes correlated  
with both explicit attitudes and judgment for feelings-based theorists, but not for reasons-based  
theorists.

**10:30**  
**Values and Indirect Attitude Change: Undermining a Value Decreases  
Favorability of Related Attitudes**  
KEVIN L. BLANKENSHIP, Iowa State University; DUANE T. WEGENER,  
Purdue University  
kevinb1@iastate.edu  
The current research examined values as an indirect route for attitude change. Specifically,  
when the favorability of a value was undermined, attitudes related to that value also became  
less favorable, compared to attitudes unrelated to the value. Thus, attitude change was observed  
without directly addressing the attitude topic at all.

**FIGURE 3-3** Sample page from the 2009 Midwestern Psychological Association meeting program.

of the program. By reading the abstracts of the papers, you can identify those papers of interest and glean something of the findings. If you want more information, you can then write or call the author. Some professional organizations now provide full programs online. Visit one or more of the Web sites for these organizations (e.g., Eastern Psychological Association, Midwestern Psychological Association, and the Society for Psychological Science) to see if online versions of programs are available.

**Other Sources of Research Information** Personal replies to your inquiries fall under the heading of **personal communications** and are yet another source of research information. Projects completed under the auspices of a grant or an agency often result in the production of a *technical report*, which can be obtained through the agency.

In addition, dissertations and theses completed by graduate students as part of their degree requirements are placed on file in the libraries of the university at which the work was done. You can find abstracts describing these studies in *Dissertation Abstracts International*, a reference work found in most college libraries. For a fee, the abstracting service will send you a copy of the complete manuscript on paper or microfilm. A Web-based service available at subscribing libraries allows you to search for relevant dissertations in this database and read their summaries.

The Internet provides yet another source of research information. You can find journal articles, technical reports, original papers, and so on via an online search. For example, entering the keyword “helping behavior” in the Google search engine turned up several hits, some of which are reports of studies done on helping behavior and altruism. Such sources, although they may prove valuable when developing ideas for research, should be used with caution because they may not be refereed. However, the Internet also provides electronic versions of refereed professional journals. For example, *The Canadian Journal of Behavioral Science* provides an online electronic version of full articles. You can even find hundreds of classic historical articles and books at the Classics in the History of Psychology Web site (at the time of this writing <http://psychclassics.yorku.ca/>). Articles and books are indexed both by author and by subject.

When judging the quality of the material you find on the Internet, use the same criteria discussed earlier (refereed versus nonrefereed, ISI ranking). You can also consult a number of online resources to help you evaluate Internet materials. For example, the Purdue University Owl Web site suggests evaluating Internet sources according to four categories: authorship (e.g., Is an author clearly indicated?), accuracy of the information (e.g., Is the information current and are sources provided?), the goals of the Web site (e.g., Is it informative, persuasive or intended to advertise?), and access (e.g., How did you find the site and are there links to other reputable sites?). Use caution if you cannot determine the quality of a resource found on the Internet.

The Internet also offers services that will allow you to search for and obtain full-text versions of articles from a variety of publications (some scholarly and some not). One such service is *Academic Search Premier* provided by EBSCOhost, which indexes articles in a variety of publications from 1990 to the present (depending on the journal). You can search for literature by subject, journal, and a host of other categories. You also can limit your search to full-text articles from peer-reviewed journals. For example, a search for full-text articles in scholarly journals on “helping behavior” (used as the keyword for the search) turned up hundreds of articles. Of course, many of the articles identified in such a search may not contain what you are looking for. You can specify additional criteria to further narrow your search. For example, replacing “altruism” with “altruism and empathy” and “personality” reduced the number of articles found to 15. Once you have located the full-text articles that interest you, you can read them online and, if you wish, print them out.

You can gain access to EBSCOhost in a couple of ways. Check with your university or local public library to determine whether it has a subscription to the service. Some states (e.g., Indiana) have contracts with EBSCOhost so that any resident of the state can access the databases for free. If you are not given free access, you can subscribe individually. See your librarian for information on subscribing to EBSCOhost.

## QUESTIONS TO PONDER

---

1. How can professional conferences provide you with information about research?
2. How can Internet resources be used to track down research information?
3. How do you assess the quality of information found on the Internet?

## PERFORMING LIBRARY RESEARCH

---

With so many sources of research information to choose from, you may find yourself quickly overwhelmed if you do not adopt an efficient strategy for separating the useful articles from less useful ones. You need a method that quickly identifies articles relevant to your topic. Ideally, the method should identify all such articles because the one that you miss may be the one that duplicates exactly what you were planning to do. Fortunately, such a method exists.

### The Basic Strategy

Although a number of variations exist, the basic strategy is this: (1) Find a relevant research article (you can do this by consulting the reference sections of textbooks or other books or tracking down an article by using a computerized database); (2) use the reference section of the article that you found to locate other articles (inspecting the titles of articles can give you some insight into the terminology used by researchers in an area); (3) repeat steps 1 and 2 for each relevant article identified until you can find no more; (4) use one of the many indexes available in your library (discussed in the next sections) to identify more recent articles; and (5) repeat the entire process as you find more and more recent articles.

**Research Tools** The most fundamental library research tool for doing a literature search is an index or a searchable electronic database. Many libraries now subscribe to a number of electronic databases that allow you to search for information sources quickly and easily. One such database is **PsycINFO**. PsycINFO includes over 1,800 journals in 25 languages, as well as books, conference papers, and dissertations. The database covers materials published as far back as 1872 through the present. PsycINFO can help you to find material that is relevant to your research topic but does not provide access to that material. One resource that does is **PsycARTICLES**, which provides online access to journals published by the APA. Through PsycARTICLES you can obtain full-text copies of articles published in APA journals. (Some libraries may integrate a number of databases under one search tool [e.g., PsycARTICLES may be integrated within Ebsco] Check with your library's database system or consult a librarian to find out what is available in your library.)

In addition to PsycINFO and PsycARTICLES, there are other electronic and hard-copy databases you can use. In the following sections, we explore some of the indexes and databases available to you. Although we can offer some basic information

on how to use these sources, the best way to learn is through hands-on experience. Also, because space limitations prohibit an in-depth exploration of all resources, we focus on using PsycINFO to do literature searches.

## QUESTIONS TO PONDER

---

1. What is the basic strategy you should follow when doing a literature search?
2. In what ways does PsycARTICLES differ from PsycINFO?

## Using PsycINFO

In the past, a student searched for articles in psychological journals using the hard-bound volumes of the *Psychological Abstracts*. The process involved scouring printed indexes to find relevant entries, searching abstracts in another volume, and finally finding the full article in the printed journals housed in the library. This process was long, laborious, and fatiguing. Fortunately, you will be spared this torturous process. Today, much of this tedious work is done by a computerized database such as PsycINFO, which allows you to search for articles, books, and book chapters rapidly and efficiently.

**Conducting a PsycINFO Search** When conducting a computer search using PsycINFO, you enter a keyword or keywords, the computer finds every instance in which those terms are used in citations contained in the PsycINFO database, and it adds those citations to your reference list. There are two ways to do this. The default search mode is a “Quick Search” that allows you to enter one or more keywords as a single entry and specify a rough date range for your search. You can also perform an “Advanced Search” (by clicking on this tab on the screen) that allows you to search three separate fields of your choosing (e.g., keyword, author, and abstract). In the advanced search mode, you can also specify a more precise range of dates. So, for example, if you want to limit your search to sources from the past 10 years, you can specify this date range. You can also specify whether you want to limit your search to journal articles only or include books as well. There is a host of other parameters that you can set to focus your search.

After you enter your search term(s), the screen displays a list of entries found. Before perusing this list, you can narrow your search via a series of tabs at the top of the entry listing. These allow you to view only peer-reviewed journal articles, books, conference presentations, or all entries found. If you are interested only in peer-reviewed (referred) journals, click on that tab.

Each record displayed on your screen includes the title, bibliographic information, and a brief description of the study. Clicking on the hyperlinked title (or on the “View Record” link) will take you to the more complete record. There is also a hyperlink labeled “References.” If you click on this link, you will see a list of the references cited in the paper that is the subject of the record. From there you can obtain the abstract of any of those references (by clicking on the “Abstract” hyperlink) or

a list of records that have cited any of those references (by clicking on the “Cited by” hyperlink). Finally, on the right side of the short record is a hyperlinked list of “Descriptors.” Clicking on one of these descriptors brings up a list of records that are indexed according to that descriptor.

**Narrowing Your Search** PsycINFO can save you a great deal of time by doing much of the tedious work of searching indexes for you. However, you may find that your search yields more citations than you can possibly look at. For example, using the keyword “stereotype threat” produced over 2,000 records. Many may not be relevant to your interests. You probably don’t want to wade through more than 2,000 abstracts to find the few that fit your needs. You must find a way to narrow your search.

One way to narrow your search is to use the check boxes in the “Descriptors” section of the full display of a PsycINFO record. You could select two of the descriptor terms that closely match the topics you are looking for and conduct a new search using those terms. Another way to narrow your search is to use the advanced search function to enter more than one keyword. In our example, you are interested in how gender relates to stereotype threat; you could enter the keyword “gender” along with “stereotype threat.” The default advanced search will search the entire PsycINFO record for records where “stereotype threat” and “gender” appear together. Doing this search yielded 832 records, still a large number. You can reduce further the number of records found by specifying that the search only look in the abstracts of PsycINFO records. This strategy reduced the number of records found to 105 (63 in peer-reviewed journals). You might also try using some of the terms in the descriptor (DE) list included on the PsycINFO record, assuming that they fit your search needs.

If you are unsure which term or terms to use in your search, consult the online version of *Thesaurus of Psychological Index Terms*. You can access the thesaurus directly from PsychINFO. The thesaurus provides information on other terms that you could use to narrow your search (by providing narrower keywords), broaden your search (by providing broader keywords), or expand your search using related terms.

**A Note of Caution About Using PsycINFO** PsycINFO and other electronic database search systems can save you a considerable amount of time and effort. However, keep in mind certain limitations of computerized search systems. A search is only as good as the keywords you enter. The computer is incredibly fast and obedient and, unfortunately, pretty stupid. It will do only what you tell it to do. It cannot think for itself and figure out what you really want when you enter a keyword. It will find every reference that includes your keywords. You may find, much to your annoyance, that terms are used more broadly in the indexed material than you anticipated. Imagine, for example, that you are looking up a topic concerning elderly individuals and decide to use the keyword “aged” (as in age-ed). You are initially excited to find more than 20,000 refereed articles using that term. Your excitement turns to irritation, however, when you discover that the majority of titles with “aged” refer to an age range (e.g., “subjects aged 12–14 years”). If this happens, use the online thesaurus to help you identify a more useful keyword.

## QUESTIONS TO PONDER

---

1. How do you perform a basic and advanced PsycINFO search?
2. How can you narrow or broaden a PsycINFO search?
3. What are the advantages and disadvantages of doing a PsycINFO search?

## Using PsycARTICLES

One disadvantage of PsycINFO is that you may not be able to obtain a full copy of an article that you want to read (check with your librarian to see if the library subscribes to full-text databases). Such is not the case with PsycARTICLES. This database comprises over 50 refereed journals published by the APA. Using the PsycARTICLES search engine, you can locate full versions of the journal articles that you want to read. For example, entering the keywords “prejudice” and “race” yielded 15 articles in several different journals. By clicking on one of the full-text options (html or pdf), you can display the full article. You then can read the article online, print the article, or save the article to disk.

The advantages of using PsycARTICLES are obvious (e.g., ease of use and access to full articles). However, there is a drawback. Your literature search using PsycARTICLES is limited to those journals published by the APA. Although these are among the top journals in psychology, the list does not include many other top-flight journals such as *Child Development*, *Personality and Social Psychology Bulletin*, or *Law and Human Behavior*. To access materials in journals not published by the APA, you would have to use another search engine.

## Other Computerized Databases

PsycINFO and PsycARTICLES are not the only electronic search resources available. Another powerful search tool is EBSCOhost. This database covers a wide range of journals in a number of areas (e.g., psychology, medicine, science, and communications and technology). You can select which database or databases you wish to search with EBSCOhost. Once you select the databases you want to use, a search page comes up. You then select “Basic Search” or “Advanced Search.” Here you can enter keywords, author names, journal titles, or article titles. You can also specify whether you want to limit your search to peer-reviewed sources and/or sources that have full-text versions.

Another search engine is IngentaConnect, which covers over 32,000 publications. Entering the keywords “stereotype threat” turned up 69 references in a number of electronic sources. IngentaConnect returns full reference citations for the articles found and access to the abstract (summary) of the article. There is, however, a charge for the full text of the article. An advantage of IngentaConnect is that you can access it directly from the Internet and do not need to go through a subscribing library. This database is a good alternative to PsycINFO if you do not have access to PsycINFO or if it is temporarily unavailable at your library.

Another computerized database you may find helpful is JSTOR, which includes journals from a wide range of fields (e.g., sociology, philosophy, anthropology, and



political science). A JSTOR search of the psychology journals in the database with the same keywords as used above uncovered 186 reference citations. JSTOR provides access to abstracts and allows you to download a full version of an article (you can limit your search to full-text sources) free of charge in a number of different formats.

You may find that JSTOR is not the best search engine for specific topics in psychology. You will not get the same kind of comprehensive results that you will with PsycINFO. However, used as a supplement to other databases, JSTOR may turn up articles that give a different perspective on your topic. This may give you a broader perspective on your topic and some ideas about research that needs to be done.

## QUESTIONS TO PONDER

---

1. What is PsycARTICLES and how can you use it to search the literature?
2. Besides PsycINFO and PsycARTICLES, what other databases can you use to search the literature?

## General Internet Resources

Reference material also can be found by using one of the many Internet search engines (e.g., Yahoo!, Google, and Ask). Entering the phrase “stereotype threat” into the Google search engine turned up a hodgepodge of links to related material. One link was to a 2004 article summarizing a talk given at a university by Claude Steele (who pioneered research on stereotype threat). Another link was to the APA *Psychology Matters* Web site, which provided an article containing an overview of research on stereotype threat.

Using a general Internet search engine can turn up a treasure trove of information. However, you must be cautious when you consider using any materials found this way on the Internet. The fact is that anyone can publish anything that he or she pleases on the Internet. Typically, materials do not undergo any kind of peer review. Consequently, you cannot be sure that the information you are getting is valid, reliable, or objective. You should read such materials with a very critical eye. As noted above, the Purdue University OWL Web site suggests that you find out about the author of the material, the affiliated institution, the timeliness of the material, the publisher (if any), the accuracy of the information, the goals of the Web site on which the information was found, and the reputation of the links that brought you to the information. The Johns Hopkins University library has an excellent document on evaluating Internet sources (<http://www.library.jhu.edu/researchhelp/general/evaluating/index.html>). Further information on evaluating Web-based information sources can be found on the Web site that accompanies this text.

## Computer Searching for Books and Other Library Materials

Many libraries have installed computerized databases indexing the books and journals housed in the library. These systems are similar to PsycINFO and allow you to search for materials by author, title, subject, and keywords. The beauty of these modern systems is that you are not limited to searching your university library. You can easily gain

**TABLE 3-3 Other Library Search Resources**

RESOURCE	APPLICATION
<i>Psychological Abstracts Subject Index</i> (hardbound version)	Used to find sources by subject matter in psychology.
<i>Psychological Abstracts Author Index</i> (hardbound version)	Used to find sources if all you have is an author's name.
<i>Social Science Citations Index: Citation Index</i>	Used when you want to find out what other, more recent articles have cited the article that you already have.
<i>Social Science Citations Index: Source Index</i>	Used to find articles when you have very little information (e.g., a citation to a study in a popular news magazine).
<i>Social Science Citations Index: Permuterm Subject Index</i>	Used to find articles by looking them up by topic. Format is similar to the <i>Psychological Abstracts</i> with a broader range of coverage.
<i>ISI Web of Science</i>	Provides access to a wide range of scientific search tools including the <i>Science Citation Index Expanded</i> , the <i>Social Science Citation Index</i> , and the <i>Arts and Humanities Citation Index</i> . The service allows access to over 10,000 journals and allows a number of flexible search strategies.

access to other library databases via the Internet. For example, using the Yahoo! search engine, type in the search term “university libraries databases.” This will take you to a list of links to university libraries that you can search. Or, you can type in the name of a specific library (e.g., Purdue University Library) to search that library’s holdings.

### Other Resources

For the most part, you will most likely be using PsycINFO and/or one of the other computerized resources we just reviewed. However, you should also be aware of some other potentially useful tools. Space does not allow a full exploration of these resources. We have summarized these in Table 3-3.

### QUESTIONS TO PONDER

1. How can you use general Internet sources to find research information and what cautions should you take before using information found on the Internet?
2. How can you search for books using Internet resources?
3. What “other” tools are available for you to perform an online literature search?

## READING A RESEARCH REPORT

---

Assume that you have obtained a copy of a research report. Knowing what you will find in it can save you time in locating specific kinds of information. The information contained in the report reflects the purposes for which it was written. These purposes include (1) arguing the need for doing the research, (2) showing how the research addresses that need, (3) clearly describing the methods used so that others can duplicate them, (4) presenting the findings of the research, and (5) integrating the findings with previous knowledge, including previous research findings and theories.

Consider the components of a typical research report and how they fulfill these purposes. Although the format of an article may vary from journal to journal, most research articles include the standard sections shown in Table 3-4. Sometimes sections are combined (e.g., results and discussion) or a section is added (e.g., design). Generally, however, research articles in psychological journals follow the outline shown in Table 3-4.

### Obtaining a Copy

After identifying relevant research reports, your next step is to obtain copies. Many libraries now subscribe to services that provide full-text articles online (e.g., through EBSCOhost, JSTOR, PsycINFO, or PsycARTICLES). You should contact your library to see which, if any, of these services are available. If these services are available, you can directly access html or pdf versions of articles on your computer. There are two caveats to this method of obtaining a research article. First, your library may not subscribe to the full-text services. Second, not all journals provide full-text access. In both of these cases, you will have to obtain your copy by using the hardbound journals stocked by your library or through interlibrary loan.

Your library has a list of all periodicals (including scientific journals) found on its shelves or stored on microfilm. Where you can find this list depends on what resources your library has. Many libraries have this information on a computerized database, perhaps linked with the general database system that you would use to search for books. Libraries without computerized systems most likely have a *Serials Index* that includes the call number assigned to each journal. Use the call number to find the

**TABLE 3-4** Parts of an APA-Style Article

Abstract
Introduction
Method
Participants or subjects
Apparatus or materials
Procedure
Results
Discussion
References

journal, just as you would to locate a book. If your library does *not* subscribe to that journal, you may still be able to obtain a copy of the article you want by submitting a request for interlibrary loan (see your librarian for advice on how to do this). Getting articles via interlibrary loan has become faster with the advent of the Internet: Articles can be faxed or e-mailed to you. However, the library may not always use these electronic methods, and in some instances it can take several days or weeks to get your article via interlibrary loan.

If you do find the article in the library, quickly scan it to determine if it is indeed relevant to your research. If so, copy the article for future reference (libraries have photocopiers available). Making a copy is legal, even if the article is copyrighted, as long as the copy is for personal use in your research. (You can find information about fair use of copyrighted material via an Internet search. One good source is at <http://www.umuc.edu/library/copy.shtml>.) Having your own copy will simplify the job of keeping track of important details. You can underline and make marginal notes right on the copy. If you become concerned about some point that you had not paid much attention to in your original reading, you can reread your copy.

### Reading the Literature Critically

When reading a journal article, think of yourself as a consumer of research. Apply the same skills to deciding whether you are going to “buy” a piece of research as you would when deciding whether to buy any other product. Critically reading and analyzing research literature (or any source of information for that matter) involves two steps: an initial appraisal and a careful analysis of content (Cornell University Library, 2000).

The *initial* appraisal involves evaluating the following (Cornell University Library, 2000):

- ♦ Author
- ♦ Date of publication
- ♦ Edition or revision
- ♦ Publisher
- ♦ Title of the journal

When evaluating the author, you should look at his or her credentials, including institutional affiliation and past experience in the area. It is important to consider the author and the author’s institutional affiliation because not all research findings are reported in scholarly journals. Some research is disseminated through “research centers” and other organizations. By evaluating the author and the institution, you can make an assessment of any potential biases. For example, a study that comes from an organization with a political agenda may not present facts in a correct or unbiased fashion. The main author of a research report from such an organization might not even be academically qualified or trained to conduct research and correctly interpret findings. One way you can check on the author is to see if the author’s work has been cited by others in the same area. Important works by respected authors are often cited by other researchers.

Look at the date of the publication to see if the source is current or potentially out of date. In some areas (e.g., neuroscience), new discoveries are made almost daily and may make older research out of date and obsolete. Try to find the most up-to-date sources that you can. When evaluating a book, determine if the copy you have is the most recent edition. If it is not, find the most recent edition because it will have been updated with the most current information available at the date of publication. Also, note the publisher for both books and journals. Some books are published by companies (sometimes called “vanity presses”) that require authors to pay for publication of their works. Books published in this way may not undergo a careful scholarly review prior to publication. Generally, books published by university publishers will be scholarly, as will books published by well-recognized publishing houses (e.g., Lawrence Erlbaum Associates). Although this is no guarantee of quality, a book from a reputable publisher will usually be of high quality. The same goes for journals. As indicated earlier, some journals are peer reviewed and some are not. Try to use peer-reviewed journals whenever possible. Finally, look at the title of the publication that you are thinking of using. This will help you determine if the publication is scholarly or not. There is no hard-and-fast rule of thumb to tell you if a publication is scholarly. Use the guidelines in Table 3-1 to determine the nature of the publication.

Evaluating the content of an article published in a scholarly psychological journal involves a careful reading and analysis of the different parts of the article (outlined in Table 3-4). In the next sections, we explore how to critically analyze each section of an APA-style journal article.

## QUESTIONS TO PONDER

---

1. Why is it important to read a research report critically?
2. What initial appraisals should you make of an article that you are going to read?

**Evaluating the Introduction** When reading the introduction to a paper, determine whether or not the author has adequately reviewed the relevant literature. Were any important papers neglected? Does the author support any assertions with reference citations? In addition, ask yourself the following:

1. Has the author correctly represented the results from previous research?  
Sometimes when authors summarize previous research, they make errors or select only findings consistent with their ideas. Also, as already noted, authors may have a theoretical orientation that may bias their summary of existing research findings. If you are suspicious, look up the original sources and evaluate them for yourself. Also, you should determine if the author has cited the most up-to-date materials. Reliance on older material may not give you an accurate picture of the current research or theory in an area.
2. Does the author clearly state the purposes of the study and the nature of the problem under study?

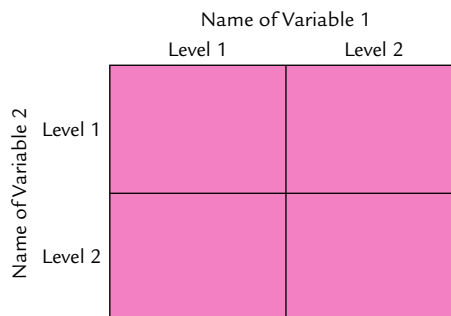
3. Do the hypotheses logically follow from the discussion in the introduction?
4. Are the hypotheses clearly stated and, more important, are they testable?

**Evaluating the Method Section** The method section describes precisely how the study was carried out. You might think of this section as a “cookbook,” or a set of directions, for conducting the study. It usually contains subsections including *participants* or *subjects* (describing the nature of the subject sample used), *materials* or *apparatus* (describing any equipment or other materials used), and *procedure* (describing precisely how the study was carried out). When reading the method section of an article, evaluate the following:

1. Who served as participants in the study? How were the participants selected (randomly, through a subject pool, etc.)? Were the participants all of one race, gender, or ethnic background? If so, this could limit the generality of the results (the degree to which the results apply beyond the parameters of the study). For example, if only male participants were used, a legitimate question is whether the results would apply as well to females. Also, look at the size of the sample. Were enough participants included to allow an adequate test of any hypotheses stated in the introduction?
2. Does the design of the study allow an adequate test of the hypotheses stated in the introduction? For example, do the levels of the independent variables allow an adequate test of the hypotheses? Is information provided about the reliability and validity of any measures used?
3. Are there any flaws in materials or procedures used that might affect the validity of the study? A good way to assess this is to map out the design (discussed next) of the study and evaluate it against the stated purpose of the study.

To better understand the design of an experiment you may want to “map out” the study. You can do this by drawing a grid or grids representing the design. For example, if you were reading about an experiment on stereotype threat which included two independent variables, a map of the experiment would look like the one shown in Figure 3-4. In Figure 3-4, the name of Variable 1 (e.g., Stereotype threat condition)

**FIGURE 3-4** Graphical display of an experimental design.



would go on top with the names of the two levels (e.g., Threat and No threat) underneath above each row. The name of Variable 2 (e.g., Test anxiety assessment) would go on the side next to the names of the two levels to the left of each row (e.g., Before and After). Each box, or cell, on the figure represents a unique combination of Variables 1 and 2. Of course, more complex designs would require more complex maps. However, the general strategy of mapping out designs, especially complex ones, can help you better conceptualize what was done in an experiment.

## QUESTIONS TO PONDER

---

1. What should you evaluate when reading the introduction to an APA-style paper?
2. What should you look for when evaluating the method section of an APA-style paper?

**Evaluating the Results Section** The results section presents the data of the study, usually in summary form (means, standard deviations, correlations, etc.), along with the results from any inferential statistical tests applied to the data (e.g., a *t* test or analysis of variance). When evaluating the results section, look for the following:

1. Which effects are statistically significant? Note which effects were significant and whether those effects are consistent with the hypotheses stated in the introduction.
2. Are the differences reported large or small? Look at the means (or other measures of center) being compared and note how much difference emerged. You may find that, although an effect is significant, it is small.
3. Were the appropriate statistics used?
4. Do the text, tables, and figures match? Sometimes errors occur in the preparation of tables and figures, so be sure to check for accuracy. Also, check to see if the author's description of the relationships depicted in any tables or figures matches what is shown.
5. If data are presented numerically in tables or in the text of the article, you should graph those results. This is especially important when complex relationships are reported among variables.

If statistics are not reported, determine whether the author has correctly described the relationships among the variables and has indicated how reliability was assessed.

**Evaluating the Discussion Section** In the discussion section, you will find the author's interpretations of the results reported. The discussion section usually begins with a summary of the major findings of the study, followed by the author's interpretations of the data and a synthesis of the findings with previous research and theory.

You also may find a discussion of any limitations of the study. When evaluating the discussion section, here are a few things to look for:

1. Do the author's conclusions follow from the results reported? Sometimes authors overstep the bounds of the results and draw unwarranted conclusions.
2. Does the author offer speculations concerning the results? In the discussion section, the author is free to speculate on the meaning of the results and on any applications. Carefully evaluate the discussion section and separate author speculations from conclusions supported directly by the results. Evaluate whether the author strays too far from the data when speculating about the implications of the results.
3. How well do the findings of the study mesh with previous research and existing theory? Are the results consistent with previous research, or are they unique? If the study is the only one that has found a certain effect (if other research has failed to find the effect or found just the opposite effect), be skeptical about the results.
4. Does the author point the way to directions for future research in the area? That is, does the author indicate other variables that might affect the behavior studied and suggest new studies to test the effects of these variables?

**References** The final section of an article is usually the reference section (a few articles include appendixes as well) in which the author lists all the references cited in the body of the paper. Complete references are provided. You can use these to find other research on your topic.

## QUESTIONS TO PONDER

---

1. What information should you evaluate in the results section of an APA-style paper?
2. What information should you evaluate in the discussion section of an APA-style paper?
3. What should you look for when evaluating the references in an APA-style paper?

## FACTORS AFFECTING THE QUALITY OF A SOURCE OF RESEARCH INFORMATION

---

One thing to keep in mind when selecting a source of information about a particular area of research is that not all books, journals, or convention presentations are created equal. Some sources of information publish original research whereas others may only summarize the findings of a study. The criteria that journals use for accepting a



manuscript determine which manuscripts will be accepted or rejected for publication, leading potentially to a bias in the content of the journal. Additionally, although most publications use a peer-review process to ensure the quality of the works published, some do not. In this section, we explore these issues and show how they relate to your literature review.

### Publication Practices

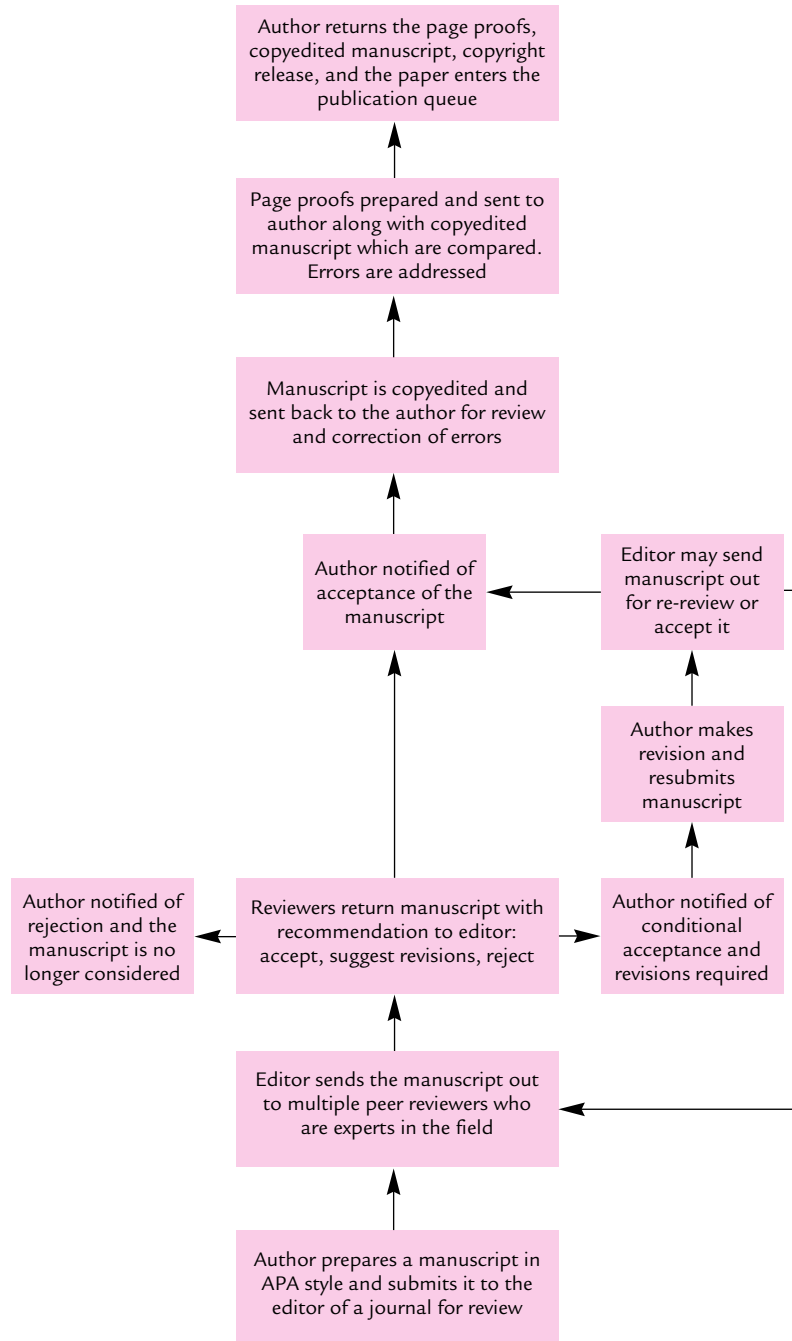
When you conduct a literature review, one question should come to mind in considering a research area as a whole: Do the articles that you are reading provide a fair and accurate picture of the state of the research in the field? Figure 3-5 shows the general process that a manuscript undergoes when it is submitted for publication. Although it is true that journals generally provide a good comprehensive view of the research within their scope, there may be research that never makes it into the hallowed pages of scientific journals because of the publication practices adopted by scholarly journals.

When a manuscript is submitted for consideration to a scholarly journal, editors and reviewers guide their evaluations of the manuscript by a set of largely unwritten rules. These include whether the results reported meet conventionally accepted levels of statistical significance, whether the findings are consistent with other findings in the area, and whether the contribution of the research to the area is significant. The policies adopted by the current editor also could affect the chances of a manuscript being accepted for publication. We examine these publication practices and their possible effects on the published literature next.

### Statistical Significance

Data collected in psychological research are usually subjected to a statistical analysis in order to determine the probability that chance and chance alone would have resulted in effects as large as or larger than those actually observed. If this probability is sufficiently low (e.g., less than .05, or 1 chance in 20), it is deemed unlikely that chance alone was responsible for the observed effect, and the effect is said to be *statistically significant*. (See Chapter 14 for a more detailed discussion of statistical significance testing.) The criterion probability used to determine statistical significance, called *alpha*, determines how often effects that are just chance differences end up being declared statistically significant. Thus, if alpha is .05, the studied effect will happen, on average, 5 times in 100 tests. In most journals, editors are reluctant to accept papers in which results fail to achieve the accepted minimum alpha level of .05. The reason, of course, is that such results stand a relatively high chance of being due to random factors rather than to the variable whose possible effect was being assessed in the study. Researchers are aware of the requirement for statistical significance and therefore usually do not report the results of studies that fail to meet it.

If the investigator is convinced that an effect is there, despite the lack of statistical significance, he or she may elect to repeat the study while using better controls or different parameters. Nothing is inherently wrong with such a strategy. If the effect is there, better control over extraneous variables and selection of more favorable parameters are likely to reveal it. If the effect is *not* there, however, repeated attempts



**FIGURE 3-5** Diagram of the editorial review process.

to demonstrate the effect eventually lead to obtaining statistically significant results by chance. Through probability pyramiding (see Chapter 14), the likelihood that this will happen is much greater than the stated alpha level would suggest.

The failures to obtain significant results generally wind up in someone's file drawer, forgotten and buried. In most cases, only those attempts that were successful in obtaining significant results are submitted for publication. Yet, because of probability pyramiding, the published results are more likely to have been significant because of chance than the stated alpha would lead us to believe. This effect is known as the **file drawer phenomenon** (Rosenthal, 1979, 1984). To the extent that the file drawer phenomenon operates, published findings as a group may be less reliable than they seem.

The problem of negative findings is serious. The failure to obtain an expected relationship can be as important for understanding and for advancement of theory as confirmation. Yet this information is difficult to disseminate to the scientific community. Laboratories may independently and needlessly duplicate each other's negative findings simply because they are unaware of each other's negative results.

## QUESTIONS TO PONDER

---

1. How do publication practices affect the articles that ultimately get published in journals?
2. What role does statistical significance play in determining what gets published in a journal?
3. What is the file drawer phenomenon and how does it relate to published research?

## Consistency With Previous Knowledge

Another criterion used to assess a research paper's acceptability is the consistency of its findings with previous knowledge. Most findings are expected to build on the existing structure of knowledge in the field, that is, to add new information, to demonstrate the applicability of known principles in new areas, and to show the limits of conditions within which a phenomenon holds. Findings that do not make sense within the currently accepted framework are suspect.

When the currently accepted framework has deep support, then such anomalous findings call into question the study that generated them rather than the framework itself. Reviewers and editors are likely to give the paper an especially critical appraisal in an attempt to identify faults in the logic and implementation of the design that may have led to the anomalous results. Ultimately, some reason may be found for rejecting the paper.

An excellent example in which this process operated was the initial work by Garcia and Koelling (1966) on learned taste aversions. Garcia and Koelling exposed thirsty rats to a solution of water that had been given a flavor unfamiliar to the rats. Some of the rats were then injected with lithium chloride, and the rest of the group was given a placebo injection of saline solution. The rats injected with lithium

chloride became ill from the injection about 6 hours later. The rats were allowed to recover and then were given a choice between drinking plain water or the flavored water. Rats injected with the saline solution showed no preference between the two, but rats injected with lithium chloride avoided the novel flavor.

From this evidence, Garcia and Koelling (1966) concluded that the rats injected with lithium chloride had formed, in a single “trial,” an association between the novel flavor and the illness. In other words, classical conditioning had occurred between a conditioned stimulus (the flavor) and an unconditioned stimulus (the illness) across a 6-hour interstimulus interval.

This was a striking finding. Classical conditioning had been extensively researched by Pavlov and others. It was well known that interstimulus intervals beyond a few *minutes* were completely ineffective in establishing a conditioned response, even when hundreds of trials were conducted. To reviewers and editors looking at Garcia and Koelling’s (1966) manuscript, something was fishy. Garcia and Koelling’s finding was a fluke, or some unreported aspect of methodology was introducing a confounding factor. The results simply couldn’t be correct. The paper was repeatedly rejected by reviewers.

It was not until others heard of Garcia and Koelling’s (1966) findings “through the grapevine” and successfully replicated their results that the phenomenon of learned taste aversions gained credibility among reviewers. Only then did papers on the topic begin to be accepted in the established refereed journals. Once accepted, Garcia and Koelling’s discovery and other similarly anomalous findings became the basis for new theories concerning the nature and limits of laws of learning (such as Seligman & Hager, 1972). Hence, in refusing to publish Garcia and Koelling’s findings, reviewers and editors delayed progress, but ultimately the new findings surfaced to challenge established thinking.

Editors and reviewers are thus in a tough position. To function effectively, they must be conservative in accepting papers that report anomalous findings. Yet they must be open-minded enough to avoid simply assuming that such findings *must* result from methodological flaws. Later in this chapter, we examine just how successful editors and reviewers have been at maintaining this balance.

### Significance of the Contribution

When determining whether to accept or reject a paper for publication, editors and reviewers must assess the degree to which the findings described in the paper contribute to the advancement of knowledge. At one time, papers were considered acceptable even if they reported only a single experiment involving simply an experimental and a control group. A researcher could publish a number of papers in a relatively short time, but each contributed little new information.

Today, journals usually insist that a paper report a series of experiments or at least a parametric study involving several levels of two or more variables. For example, a paper might report a first experiment that demonstrates a relationship between two variables. Several follow-up experiments might then appear that trace the effective range of the independent variable and test various alternative explanations for the relationship. Such a paper provides a fair amount of information about the

phenomenon under investigation and, in pursuing the phenomenon through several experiments, also demonstrates the phenomenon's reliability through immediate systematic replication.

Although these are important advantages, insisting on multiple experiments or studies within a paper also can have a negative side. Although the study provides more information, the information contained in the study cannot see the light of day until the entire series of experiments or observations has been completed. The resulting paper is more time-consuming to review and evaluate. Reviewers have more opportunities to find defects that may require modification of the manuscript. The result is delay in getting out what may be an important finding to the scientific community.

### Editorial Policy

Editorial policy is yet another factor that can influence what appears in journals. Frequently, an area of research becomes “hot,” resulting in a flood of articles in the area. Researchers latch on to a particular research area (e.g., eyewitness identification, day care in early infancy) and investigate it, sometimes to the exclusion of other important research areas. When this happens, a journal editor may take steps to ensure that more variety appears in a journal. For example, research on eyewitness identification has been a hot topic for the past several years. Interest reached its peak in the 1980s, and the premier journal in the area, *Law and Human Behavior*, published a large number of articles on this topic—perhaps too large a number. In 1986 Michael Saks took over as editor of the journal. He made it clear that he was going to give preference to manuscripts dealing with issues other than eyewitness identification.

Editorial policy also can show itself if the editor enters an unintended bias into the review system. The editor is the one who decides whether a paper will be sent out for review and, ultimately, if it will be published. If the editor has a bias—say, toward a particular theory—that editor may be unwilling to publish articles that do not support that theory. We discuss this issue in the next section.

### QUESTIONS TO PONDER

---

1. How can consistency with previous knowledge affect whether a paper gets published in a journal?
2. How does the significance of a contribution influence an editor's decision to publish a paper in a journal?
3. How can editorial policy affect whether a paper gets published in a journal?

### Peer Review

Some sources of information (including books, journal articles, and convention proceedings) use a **peer-review** process. This means that the materials to be published or presented are reviewed by experts in the area that the material covers. These experts will receive a copy of the materials and do a thorough review of the content. They

will then recommend to the editor of the journal whether to accept (either outright or with revisions) or reject the manuscript.

Although peer review is a time-honored tradition in science as a way to ensure quality, it is far from perfect. Just because something is published in a refereed journal does not mean that it is a sound or important piece of research. Conversely, you may find some gems in nonrefereed journals. The reason for this seeming lack of consistency has to do with problems in the peer-review process.

**Problems With Peer Review** As we noted, when you send a manuscript to a refereed journal, the editor will send your manuscript to expert reviewers (usually two) who will carefully evaluate your paper. In some cases, peer review is anonymous, and in others it is not. Individual journals determine their peer-review policies, and some choose to use anonymous peer review. Anonymous peer review might be necessary, but it does have its problems. Although you hope that your colleagues in research are honest and fair in their appraisals of your work, someone with a personal dislike for you or your ideas could sabotage your efforts. Even in the absence of malice, the reviewer may judge your manuscript unfairly because of a lack of knowledge, a bias against your general approach to research, or misreading.

Suls and Martin (2009) suggest that there are several problems with traditional peer review in the social sciences that may make the process unfair or biased. Suls and Martin point out that even though reviewers are supposed to be “experts,” reviewers may lack expertise relating to the methods and issues within a field. Even if a reviewer is a true expert, he or she may be a direct competitor of the author of an article, coloring that reviewer’s evaluation. Suls and Martin also suggest that using anonymous reviewers might encourage reviewers to be overly critical of a paper due to lack of accountability. Reviewers may also see their role as “gatekeeper” and be overly harsh and critical especially for articles with controversial content. Suls and Martin also note that a frequent criticism of peer review is that papers from well-established authors are treated more leniently than papers from lesser-known authors. The result is that papers from well-known authors have a better chance of being published, even if the work is of lower quality. Yet another problem with peer review is low levels of agreement between reviewers of the same paper (Suls & Martin, 2009).

The extent to which such factors operate within the peer-review system has been the subject of research and debate over the past three decades. For example, Mahoney (1977) investigated the influence of several properties of a research manuscript on its acceptance for publication by reviewers. With the approval of the editor, Mahoney sent manuscripts describing a fictitious experiment to 75 reviewers working for the *Journal of Applied Behavior Analysis*. Five randomly constituted groups of reviewers received different versions of the manuscript that varied according to their results and interpretations of those results. Mahoney found that the paper was consistently rated higher if its results supported the reviewer’s theoretical bias and lower if they did not. How the results were interpreted had little impact. Similarly, the recommendation to accept or reject the paper for publication was strongly influenced by the direction of the data. If the data supported the reviewer’s theoretical leanings, the reviewer usually recommended acceptance. If the data argued against those leanings, the reviewer usually recommended rejection or major revision.

Mahoney's (1977) findings showed that results favorable or unfavorable to the reviewer's point of view affect how the reviewer receives the manuscript. If the results are favorable, the reviewer is likely to believe that the results are valid and that the methodology was adequate. If the results are unfavorable, however, the reviewer is likely to believe that the study must be defective. The reviewer will search diligently for flaws in the design or execution of the study and use even minor problems as reasons for rejection.

Partly because of such sources of bias, estimates of inter-reviewer reliability in the social sciences have tended to be low. Fiske and Fogg (1990) examined 402 reviews of 153 papers and found almost no agreement among reviewers, not because the reviewers overtly disagreed but because the reviewers found different aspects of the papers to criticize. It was as if they had read different papers! Lindsey (1978), in his book *The Scientific Publication System in Social Science*, notes that empirical studies have consistently found reliabilities of around .25 (the correlation between reviewer judgments). Whether both reviewers will agree that your paper is publishable is thus very nearly a chance affair.

The unreliability of the peer-review system was highlighted by a study conducted by Peters and Ceci (1982). Peters and Ceci identified 12 *published* articles that had appeared in different major psychology journals. Each article was authored by at least one individual from a "prestige" institution and had appeared between 18 and 32 months earlier. The names of the original authors and their institutional affiliations were removed and replaced by fictitious names and affiliations. In addition, the titles, abstracts, and introductions were cosmetically altered (without changing the content) to reduce the chances that the articles would be recognized. Retyped as manuscripts, the articles were then resubmitted to *the same journals that had originally published them* (and in most cases, to the same editor).

The results were dramatic. Only 3 of the 12 articles were identified as resubmissions and rejected for this reason. The remaining 9 were undetected. Of those 9, 8 *were rejected for publication*. Even more amazing, in every case both reviewers agreed and the editor concurred.

Because the articles had appeared before, the reviewers might have rejected the papers because they remembered the earlier data (although not the articles themselves) and thus viewed the information they contained as contributing nothing new. If this were the case, however, no hint of this was given in the reasons cited by the reviewers. According to Peters and Ceci (1982), the reasons given for rejecting the papers usually concerned major flaws in the methodology. Thus, papers that had already been accepted into the archival literature only months earlier were subsequently seen as too methodologically flawed to merit publication.

Peters and Ceci (1982) offer two possible reasons for the new attitude toward the papers. The change in authorship and affiliation from prestigious to unknown may have had a negative influence on the evaluation. Or, because of the approximately 80% rejection rate, peer review may have been so unreliable that the chances of getting positive evaluations were just too low to expect acceptance the second time. This latter view assumes that getting a positive evaluation is essentially a matter of chance for manuscripts that cannot be rejected out of hand for obvious fatal flaws. Whether either or both possibilities are true, the implication is that acceptance of

your paper (given that it is reasonably good) depends to a large extent on factors that are not under your control.

Despite the problems associated with the peer-review process, it does work pretty well. Although peer review is no guarantee that all things published are of impeccable quality, it does provide a measure of confidence that what you read is valid and reliable.

## QUESTIONS TO PONDER

---

1. What is peer review and what are the major problems associated with the practice?
2. How can the peer-review process affect the likelihood that a paper will be published in a journal?
3. What evidence is there that the peer-review process affects publication practices?

## Values Reflected in Research

Another thing you need to take into account when evaluating research is whether an author's values and beliefs have influenced the research hypotheses tested and how results are interpreted. Scientists are human beings and have their own attitudes, values, and biases. These sometimes show up in published research. The validity of research may be reduced inadvertently by allowing general cultural values, political agendas, and personal values of the researcher to influence the research process. Although we would like to think of research as "value free" and objective, some philosophers of science suggest that research cannot easily be separated from a set of values dominating a culture or a person (Longino, 1990).

Values can influence the course of scientific inquiry in several ways. Helen Longino (1990), for example, lists five nonmutually exclusive categories (p. 86):

1. *Practices*. Values can influence the practice of science, which affects the integrity of the knowledge gained by the scientific method.
2. *Questions*. Values can determine which questions are addressed and which are ignored about a given phenomenon.
3. *Data*. Values can affect how data are interpreted. Value-laden terms can affect how research data are described. Values also can determine which data are selected for analysis and the decision concerning which phenomena are to be studied in the first place.
4. *Specific assumptions*. Values influence the basic assumptions that scientists make concerning the phenomena that they study. This may cause a scientist to make inferences in specific areas of study.
5. *Global assumptions*. Values can affect the nature of the global assumptions that scientists make that can affect the nature and character of the research conducted in an entire area.



Similarly, David Myers (1999) indicates three broad areas that combine some of Longino's categories. First, values can affect the topics that scientists choose to study and how they are studied. Second, values can affect how we interpret observations that we make and results that we uncover. Third, values can come into play when research findings are translated into statements of what "ought to be."

**How Values Influence What and How Scientists Study** Cultural values can be seen operating on science. For example, in the United States, researchers interested in conformity effects have focused on the role of the majority in influencing the minority. This probably filters down from the American political system in which the "majority rules." In Europe where there are parliamentary democracies, minority viewpoints are often taken into account when political coalitions are formed. As a consequence, much of the research on how a minority can influence a majority came out of European psychological laboratories.

Even within a culture, values can influence what we study. For example, feminist scholars point out that assumptions about gender can influence how research questions are formulated (Unger & Crawford, 1992). For example, research on the effects of early infant day care on children is usually couched in negative terms concerning how maternal employment may adversely affect a child's development. Rarely are such questions phrased in terms of the potential positive outcomes (Unger & Crawford, 1992).

Unger and Crawford (1992) also point out that gender may play a role in the manner in which research hypotheses are tested. They suggest, for example, that focusing on quantitative data (representing behavior with numerical values) may be biased against female research participants. They suggest that research also should be done that focuses on qualitative data. Such a focus would lead to a richer understanding of the motives underlying behavior. They also point out that research designs are not value neutral. Overreliance on rigid, laboratory experimentation, according to Unger and Crawford, divorces social behavior from its normal social context. They suggest using more field-oriented techniques. They do not advocate, however, abandoning experimental techniques.

**Interpreting Behavior** Scientists do not merely "read" what is out there in nature. Rather, scientists interpret what they observe (Myers, 1999). One's personal biases and cultural values may exert a strong influence over how a particular behavior is interpreted. For example, a scientist who harbors a prejudice against Blacks may be more likely to label a Black child's behavior as more aggressive than the same behavior committed by a White child. A conservative scientist may favor a biological explanation for aggression whereas a more liberal one may favor a societal explanation. In both cases, the values of the researcher provide an overarching view of the world that biases his or her interpretations of a behavioral event.

**Moving From What Is to What Ought to Be** Values also can creep into science when scientists go beyond describing and explaining relationships and begin to speculate on what ought to be (Myers, 1999). That is, scientists allow values to creep into the research process when they endeavor to define what is "right" or "normal" based on research findings. On another level, this influence of values also is seen when

researchers conduct research to influence the course of political and social events. Some feminist scholars, for example, suggest that we should not only acknowledge that values enter into science but also use them to evaluate all aspects of the research process (Unger, 1983). According to this view, science should be used to foster social change and challenge existing power structures (Peplau & Conrad, 1989).

Making sense of research requires that you be aware of the biases and other sources of error that afflict research. Given the ubiquitous nature of these sources, it is not surprising that research findings within a given area often appear contradictory.

## DEVELOPING HYPOTHESES

---

All the library research and critical reading that you have done has now put you on the threshold of the next major step in the research process: developing your idea into a testable hypothesis. This hypothesis, as we pointed out in Chapter 1, will be a tentative statement relating two (or more) variables that you are interested in studying. Your hypothesis should flow logically from the sources of information used to develop your research question. That is, given what you already know from previous research (either your own or what you read in the journals), you should be able to make a tentative statement about how your variables of interest relate to one another.

Hypothesis development is an important step in the research process because it will drive your later decisions concerning the variables to be manipulated and measured in your study. Because a poorly conceptualized research hypothesis may lead to invalid results, take considerable care when stating your hypothesis.

As an example, imagine that your general research question centers on the relationship between aging and memory. You have spent several hours in the library using PsycINFO to find relevant research articles. You have found several articles showing that older adults show poorer memory performance on tasks such as learning nonsense syllables, learning lists of words, and recognizing pictures. However, you find very little on age differences in the ability to recall details of a complex event such as a crime. Based on what you found about age differences in memory from your literature review, you strongly suspect that older adults will not recall the details of a complex event as well as younger adults.

Thus far, you have a general research question that centers on age differences in the ability to recall details of a complex event. You have identified two variables to study: participant age and memory for a complex event. Your next step is to translate your suspicion about the relationship between these two variables into a testable research hypothesis. You might, for example, develop the following hypothesis:

Older adults are expected to recall fewer details of a complex event correctly than younger adults.

Notice that you have taken the two variables from your general research question and have linked them with a specific statement concerning the expected relationship between them. This is the essence of distilling a general research question into a testable hypothesis.

Once you have developed your hypothesis, your next task is to decide how to test it. You must make a variety of important decisions concerning how to conduct your study. The next chapter explores the major issues you will face during the preliminary stages of planning your study.

## QUESTIONS TO PONDER

---

1. How do values affect the research process?
2. How do you develop hypotheses for research?

## SUMMARY

---

Sources of research ideas include experience (unsystematic observation and systematic observation), theory, and the need to solve a practical problem. Unsystematic observation includes casual observation of both human and animal behavior. Systematic observation includes carefully planned personal observations, published research reports, and your own previous or ongoing research. Theory is a set of assumptions about the causes of a phenomenon and the rules that specify how causes act; predictions made by theory can provide testable research hypotheses. The need to solve a real-world, applied issue can also be a rich source of research ideas.

Developing good research questions begins by asking questions that are answerable through objective observations. Such questions are said to be empirical. Before a question can be answered through objective observation, its terms must be supplied with operational definitions. An operational definition defines a variable in terms of the operations required to measure it. Operationally defined variables can be measured precisely but may lack generality. You must also strive to answer the “right” questions. There are some questions (e.g., “Is abortion moral or immoral?”) that are not addressable with scientific, empirical observation. You must develop questions that lend themselves to scientific scrutiny.

Good research questions should address important issues. A research question is probably important if (1) answering it will clarify relationships among variables known to affect the behavioral system under study, (2) the answer can support only one of several competing hypotheses, or (3) the answer leads to obvious practical applications. A research question is probably unimportant if (1) its answer is already firmly established, (2) the variables under scrutiny are known to have small, theoretically uninteresting effects, or (3) there is no a priori reason to believe the variables in question are causally related.

After developing your research idea, you need to conduct a careful review of the literature in your area of interest. Conducting a careful literature review can prevent you from carrying out a study that has already been done, can identify questions that need to be answered, and can help you get ideas for designing your study.

Research information can be found in several different types of sources. The best source of information is a scholarly source, such as a journal. This type of source

centers on research and theory in a given area. Another source is a substantial publication containing information that rests on a solid base of research findings. A popular publication, intended for the general population, may have articles relevant to your topic of study. However, you will not find original sources or reference citations in these publications. Finally, a sensational publication is intended to arouse curiosity or emotion. Typically, information from such a source cannot be trusted.

Scholarly information can be found in books and journals, at conventions and meetings of professional associations, and through other sources such as the Internet. Books come in a variety of forms, including original works, anthologies, and textbooks. Some books contain original material whereas others have secondhand material. Books are a useful source of information, but they may be unreviewed or out of date. Scholarly journals provide articles on current theory and research. Some journals are refereed (the articles undergo prepublication peer review) whereas others are nonrefereed (there is no prepublication review). Generally, articles in refereed journals are preferred over those in nonrefereed journals. The most up-to-date information is presented at professional conventions and meetings. You also can find research information on the Internet through sources such as EBSCOhost.

The basic strategy to follow in reviewing the literature is (1) find a relevant review article, (2) use the references in the article to find other relevant articles, and (3) use one of the many literature search tools to locate more recent articles. A number of research tools are available to help you, including PsycINFO, PsycARTICLES, EBSCOhost, and Ingenta.

Research reports follow a standard format that includes an abstract, introduction, method section, results section, discussion section, and references. Each section has a specific purpose. When you read a research report, read it critically, asking questions about the soundness of the reasoning in the introduction, the adequacy of the methods to test the hypothesis, and how well the data were analyzed and interpreted. A good rule of thumb to follow when reading critically is to be skeptical of everything you read.

Publication practices are one source of bias in scientific findings. Criteria for publication of a manuscript in a scientific journal include statistical significance of the results, consistency of results with previous findings, and editorial policy. Each of these can affect which manuscripts are eventually accepted for publication. The result is that published articles are only those that meet subjective, and somewhat strict, publication criteria.

The peer-review process is intended to ensure the quality of the product in scientific journals. Peer review involves an editor of a journal sending your manuscript to two (perhaps more) experts in your research field. The reviewers are expected to read your work and pass judgment. Unfortunately, peer reviewers are affected by personal bias. For example, reviewers are more likely to find fault with a manuscript if the reported results do not agree with their personal views on the issue studied.

Values also can enter the research process and affect the process of research, the types of questions asked, how data are interpreted, and the types of assumptions made about phenomena under study. Values also can enter into science when scientists translate their findings into what “ought to be.” Although many scientists believe that values should not be allowed to creep into research, others believe that values should be acknowledged and used to evaluate all aspects of science.

Your literature review and careful, critical reading of the sources that you found lead you to the next step in the research process: developing a testable hypothesis. A hypothesis is a tentative statement relating two (or more) variables that you are interested in studying. This is an important step in your research because your hypothesis will influence later decisions about which variables to measure or manipulate. Developing a research hypothesis involves taking your general research question and translating them into a statement that clearly specifies the expected relationships among variables.

## KEY TERMS

---

empirical question

operational definition

literature review

primary source

secondary source

refereed journal

nonrefereed journal

paper session

poster session

personal communications

PsycINFO

PsycARTICLES

*Thesaurus of Psychological Index Terms*

file drawer phenomenon

peer-review

## 4

## CHAPTER

**CHAPTER OUTLINE****Functions of a Research Design****Causal Versus Correlational Relationships****Correlational Research**

An Example of Correlational Research: Cell Phone Use and Motor Vehicle Accidents

Behavior Causation and the Correlational Approach

Why Use Correlational Research?

**Experimental Research**

Characteristics of Experimental Research

An Example of Experimental Research: Cell Phone Use While Driving

Strengths and Limitations of the Experimental Approach

Experiments Versus Demonstrations

**Internal and External Validity**

Internal Validity

External Validity

Internal Versus External Validity

**Research Settings**

The Laboratory Setting

The Field Setting

A Look Ahead

**Summary****Key Terms**

# Choosing a Research Design

After spending long hours reading and digesting the literature in a particular research area, you have isolated a behavior that needs further investigation. You have identified some potentially important variables and probably have become familiar with the methods commonly used to measure that behavior. You may even have developed some possible explanations for the relationships that you have identified through your reading and personal experience. You are now ready to choose a research design that will allow you to evaluate the relationships that you suspect exist.

Choosing an appropriate research design is crucially important to the success of your project. The decisions you make at this stage of the research process do much to determine the quality of the conclusions you can draw from your research results. This chapter identifies the problems you must face when choosing a research design, introduces the major types of research design, and describes how each type attempts to solve (or at least cope with) these problems.

## FUNCTIONS OF A RESEARCH DESIGN

Scientific studies tend to focus on one or the other of two major activities. The first activity consists of *exploratory data collection and analysis*, which is aimed at classifying behaviors within a given area of research, identifying potentially important variables, and identifying relationships between those variables and the behaviors. Such exploration is typical of the early stages of research in an area. The second activity, called *hypothesis testing*, consists of evaluating potential explanations for the observed relationships. Testable explanations allow you to predict what relationships should and should not be observed if the explanation is correct. Hypothesis testing usually begins after you have collected enough information about the behavior to begin developing supportable explanations.

## CAUSAL VERSUS CORRELATIONAL RELATIONSHIPS

---

The relationships that you identify in these activities fall into two broad categories: causal and correlational. In a **causal relationship**, one variable directly or indirectly influences another. In other words, changes in the value of one variable directly or indirectly cause changes in the value of a second. For example, if you accidentally drop a brick on your toe, the impact of the brick will probably set off a chain of events (stimulation of pain receptors in your toe, avalanche of neural impulses traveling up your leg to the spinal cord and from there to your brain, registration of pain in your brain, involuntary scream). Although there are several intervening steps between the impact of the brick on your toe and the scream, it would be proper in this case to conclude that dropping the brick on your toe *causes* you to scream. This is because it is possible to trace an unbroken chain of physical influence running from the initial event (impact of brick on toe) to the final result (scream).

Causal relationships can be unidirectional, in which case Variable A influences Variable B but not vice versa. The impact of the brick (A) may produce a scream (B), but screaming (B) does not cause the impact of the brick on your toe (A). They also can be bidirectional, in which case each variable influences the other. Everything else being equal, reducing the amount of exercise a person gets leads to weight gain. Because of the increased effort involved, heavier people tend to exercise less. Thus, exercise influences body weight, and body weight influences exercise. Even more complex causal relationships exist, and teasing them out may require considerable ingenuity on the part of the investigator. In each case, however, one can identify a set of physical influences that ties the variables together.

Simply observing that changes in one variable tend to be associated with changes in another is not enough to establish that the relationship between them is a causal one. In a **correlational relationship**, changes in one variable accompany changes in another, but the proper tests have not been conducted to show that either variable actually influences the other. Thus, all that is known is that a relationship between them exists. When changes in one variable tend to be accompanied by specific changes in another, the two variables are said to *covary*. However, such covariation does not necessarily mean that either variable exerts an influence on the other (although it may). The number of baseball games and the number of mosquitoes tend to covary (both increase in the spring and decrease in the fall), yet you would not conclude that mosquitoes cause baseball games or vice versa.

When you first begin to develop explanations for a given behavior, knowledge of observed relationships can serve as an important guide even though you may not yet know which relationships are causal. You simply make your best guess and then develop your explanation based on the causal relationships that you think exist. The validity of your explanation will then depend in part on whether the proposed causal relationships turn out, on closer examination, to be in fact causal. Distinguishing between causal and correlational relationships is thus an important part of the research process, particularly in the hypothesis-testing phase.

Your ability to identify causal relationships and to distinguish causal from correlational relationships varies with the degree of control that you have over the variables under study. The next sections describe two broad types of research design: correlational

and experimental. Both approaches allow you to identify relationships among variables, but they differ in the degree of control exerted over variables and in the ability to identify causal relationships. We begin with correlational research.

## QUESTIONS TO PONDER

---

1. How are correlational and causal relationships similar, and how are they different?
2. Can a causal relationship be bidirectional? Explain.

## CORRELATIONAL RESEARCH

---

In **correlational research**, your main interest is to determine whether two (or more) variables covary and, if so, to establish the directions, magnitudes, and forms of the observed relationships. The strategy involves developing measures of the variables of interest and collecting your data.

Correlational research belongs to a broader category called *nonexperimental* research, which also includes designs not specifically aimed at identifying relationships between variables. The latter type of research, for example, might seek to determine the average values and typical spread of scores on certain variables (e.g., grade point average and SAT scores) in a given population (e.g., applicants for admission to a particular university). Strictly speaking, such a study would be nonexperimental but not correlational. Our discussion here focuses on those nonexperimental methods used to identify and characterize relationships.

Correlational research involves observing the values of two or more variables and determining what relationships exist between them. In correlational research, you make no attempt to manipulate variables but observe them “as is.” For example, imagine that you wished to determine the nature of the relationship, if any, between pretest anxiety and test performance in introductory psychology students on campus. On test day, you have each student rate his or her own level of pretest anxiety and, after the test results are in, you determine the test performances of those same students. Your data consist of two scores for each student: self-rated anxiety level and test score. You analyze your data to determine the relationship (if any) between these variables. Note that both anxiety level and test score are simply observed as found in each student.

In some types of correlational research, you compare the average value of some variable across preformed groups of individuals where membership in a group depends on characteristics or circumstances of the participant (such as political party affiliation, eye color, handedness, occupation, economic level, or age). For example, you might compare Democrats to Republicans on attitudes toward education. Such a study would qualify as correlational research because group membership (whether Democrat or Republican) was determined by the participants’ choice of party and was not in the hands of the researcher.

Establishing that a correlational relationship exists between two variables makes it possible to predict from the value of one variable the probable value of the other



variable. For example, if you know that college grade point average (GPA) is correlated with Scholastic Assessment Test (SAT) scores, then you can use a student's SAT score to predict (within limits) the student's college GPA.

When you use correlational relationships for prediction, the variable used to predict is called the *predictor variable*, and the variable whose value is being predicted is called the *criterion variable*. Whether the linkage between these variables is causal remains an open question.

### An Example of Correlational Research: Cell Phone Use and Motor Vehicle Accidents

The opening vignette of Chapter 1 described the case of Bailey Goodman, the driver whose fatal crash may have resulted from distraction while texting on a cell phone. Even before texting became popular, researchers had already begun to investigate the possible dangers of cell phone use while driving. In 1997, David Redelmeier and Robert Tibshirani published a correlational study that examined the relationship between motor vehicle collisions and cell phone use. Drivers who had been involved in motor vehicle collisions that produced substantial property damage but no personal injury were recruited for the study. The cell phone records of these drivers were obtained for the day of the collision and for the preceding seven days. These records allowed Redelmeier and Tibshirani to compare the incidence of cell phone use during or just prior to the accident to its incidence at other times. They found that cell phone use “was associated with a quadrupling of the risk of a motor vehicle collision” (Redelmeier & Tibshirani, 1997, p. 455). McEvoy, Stevenson, McCartt, and colleagues (2005) obtained nearly identical results in a similar study involving drivers whose accidents had resulted in hospital attendance.

**Assessing the Redelmeier and Tibshirani Study** What qualifies Redelmeier and Tibshirani's study as a correlational study? In their study, cell phone usage at the time of the accident and at other times was simply recorded as found. No attempt was made to manipulate variables in order to observe any potential effects of those variables.

### Behavior Causation and the Correlational Approach

Given the results obtained by Redelmeier and Tibshirani's (1997) study and by McEvoy et al. (2005), you might be tempted to conclude that using a cell phone while driving *causes* motor vehicle accidents. However, this conclusion that a causal relationship exists is inappropriate even though the relationship appears compelling. Two obstacles stand in the way of drawing clear causal inferences from correlational data: the third-variable problem and the directionality problem.

**The Third-Variable Problem** To establish a causal relationship between two variables, you must be able to demonstrate that variation in one of the observed variables could only be due to the influence of the other observed variable. In the example, you want to show that variation in the cell phone use while driving causes variation in the risk of a motor vehicle accident. However, because the drivers (and not the researchers) chose whether or not to use a cell phone while driving, it is possible that

the observed relationship between cell phone use and the risk of a motor vehicle accident may actually be due to the influence of a third variable. For example, drivers may be more likely to talk on a cell phone while driving when they are distressed about some personal matter. This distress might also compromise a driver's ability to focus on his or her driving, thus leading to an increased risk of an accident. Although far-fetched, such a possibility cannot be ruled out in the studies cited.

The possibility that correlational relationships may result from the action of an unobserved "third variable" is called the **third-variable problem**. This unobserved variable may influence both of the observed variables (e.g., cell phone use and the likelihood of having a motor vehicle accident), causing them to vary together even though no direct relationship exists between them. The two observed variables thus may be strongly correlated even though neither variable causes changes in the other.

To resolve the third-variable problem, you must examine the effects of each potential third variable to determine whether it does, in fact, account for the observed relationship. Techniques to evaluate and statistically control the effects of such variables are available (see Chapter 15).

**The Directionality Problem** A second reason why it is hazardous to draw causal inferences from correlational data is that, even when a direct causal relationship exists, the direction of causality is sometimes difficult to determine. This difficulty is known as the **directionality problem**.

The directionality problem does not apply to the cell phone studies as it is not possible that having a motor vehicle accident could cause a person to be using a cell phone in the minutes or seconds preceding the accident. However, it can pose a problem for some studies. For example, Anderson and Dill (2000) found a positive relationship between level of aggression (as self-reported by students in their questionnaires) and the amount of exposure to violent video games. You might be tempted to conclude that students become more aggressive from playing violent video games, but it seems just as reasonable to turn the causal arrow around. Perhaps finding gratification in aggressive behavior leads to a preference for playing violent video games.

### Why Use Correlational Research?

Given the problems of interpreting the results of correlational research, you may wonder why you would want to use this approach. However, correlational research has a variety of applications, and there are many reasons to consider using it. In this section, we discuss three situations in which a correlational approach makes good sense.

**Gathering Data in the Early Stages of Research** During the initial, exploratory stage of a research project, the correlational approach's ability to identify potential causal relationships can provide a rich source of hypotheses that later may be tested experimentally. Consider the following example.

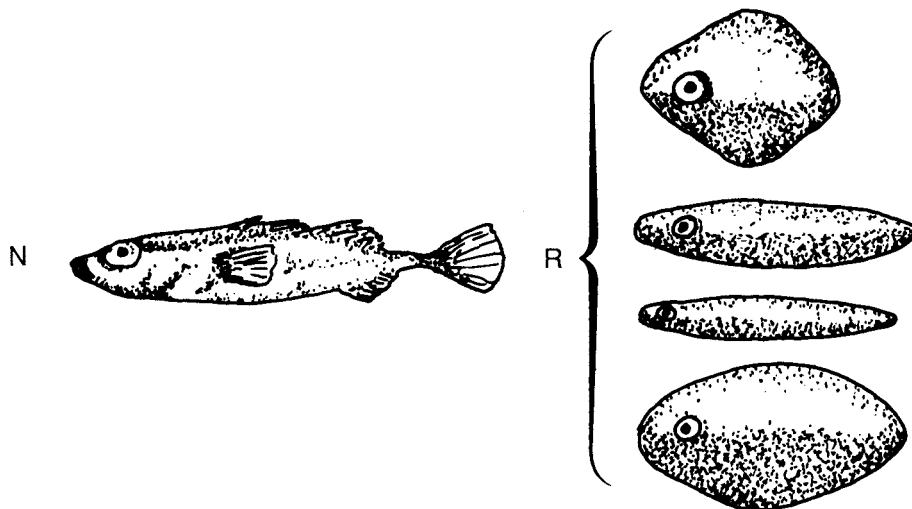
Niko Tinbergen (1951) became interested in the behavior of the three-spined stickleback, a fish that inhabits the bottoms of sandy streams in Europe. Observing sticklebacks in their natural habitat, Tinbergen found that, during the spring, the male stickleback claims a small area of a streambed and builds a cylindrically shaped nest at

its center. At the same time, the male's underbelly changes from the usual dull color to a bright red, and the male begins to drive other males from the territory surrounding the nest. Female sticklebacks lack this coloration and are not driven away by the males.

These initial observations were purely correlational and as such do not allow one to draw firm conclusions with respect to cause and effect. The observations showed that the defending male's behavior toward an intruding stickleback is correlated with the intruder's physical characteristics, but which characteristics actually determine whether or not an attack will occur? Certainly many cues, such as the male's red coloration, his shape, or even perhaps his odor, could be responsible. However, these cues always appeared and disappeared together (along with the fish to which they belonged). So there was no way, through correlational study alone, to determine whether the red coloration was the actual cause of the defensive behavior or merely an ineffective correlate.

To disentangle these variables, Tinbergen (1951) turned to the experimental approach. He set up an artificial stream in his laboratory and brought in several male sticklebacks. The fish soon adapted to the new surroundings, setting up territories and building nests. Tinbergen then constructed a number of models designed to mimic several characteristics of male sticklebacks. These models ranged from one that faithfully duplicated the appearance (but not the smell) of a real stickleback to one that was just a gray disk (Figure 4-1). Some of the models included red coloration, and some did not.

When the realistic model with a red underbelly was waved past a male stickleback in the artificial stream, the male immediately tried to drive it away. Odor obviously was not necessary to elicit defensive behavior. However, Tinbergen (1951) soon discovered that almost any model with red color elicited the response. The only requirements were that the model include an eyespot near the top and that the red color appear below the eyespot.



**FIGURE 4-1** Stimuli used by Tinbergen to follow up on initial observations made in the field: N, neutral underbelly; R, red underbelly.

SOURCE: Tinbergen, 1951; reprinted with permission.

By manipulating factors such as color and shape, Tinbergen (1951) could experimentally identify the factors that were necessary to elicit the behavior. The earlier, correlational research conducted in a naturalistic (and therefore poorly controlled) setting had paved the way for the more definitive research that followed.

***Inability to Manipulate Variables*** In an experimental design, variables are manipulated to determine their effects on other variables. A second reason for choosing a correlational design over an experimental one is that manipulating the variables of interest may be impossible or unethical (see Chapter 7 for a discussion of ethics). For example, imagine that you were interested in determining whether psychopathic personality develops when a child is raised by cold, uncaring parents. To establish a clear causal connection between the parents' behavior toward the child and psychopathic personality, you would have to conduct an experiment in which the parents' behavior was manipulated by assigning infants at random to be raised by either normal parents or cold, uncaring ones. However, this experiment would be impossible to carry out (who would allow their child to participate in such an experiment?) and, because of its potential for inflicting serious harm on the child, unethical as well. In such cases, a correlational design may be the only practical and ethical option.

***Relating Naturally Occurring Variables*** A third situation in which you may choose a correlational research design over an experimental design is one in which you want to see how naturally occurring variables relate in the real world. Such information can be used to make useful predictions even if the reasons for the discovered relationships are not clear. High school GPA, scores on the SAT, class rank, and scores on the Nelson–Denny reading comprehension test correlate well with each other and with performance in college. Knowledge of these relationships has been used to predict college success. Certain theoretical views also may lead to predictions about which real-world variables should be correlated with which. These predictions can be tested by using a correlational design.

## QUESTIONS TO PONDER

---

1. What are the defining features of correlational research?
2. Why is it inappropriate to draw causal inferences from correlational data?
3. Under what conditions is correlational research preferred over experimental research?

## EXPERIMENTAL RESEARCH

---

Unlike correlational research, experimental research incorporates a high degree of control over the variables of your study. This control, if used properly, permits you to establish causal relationships among your variables. This section describes the defining characteristics of experimental research and explains how these characteristics enable us to identify causal relationships in data.

## Characteristics of Experimental Research

**Experimental research** has two defining characteristics: manipulation of one or more independent variables and control over extraneous variables. Be sure that you understand these concepts, described as follows, because they are central to understanding experimental research.

**Manipulation of Independent Variables** An **independent variable** is a variable whose values are chosen and set by the experimenter. (Another way to look at it is that the value of the independent variable is independent of the participant's behavior.) We call these set values the *levels* of the independent variable. For example, imagine that you want to determine how sleep deprivation affects a person's ability to recall previously memorized material. To examine this relationship, you might assign participants to one of three groups defined by the number of hours of sleep deprivation: 0 hours (rested), 24 hours, and 48 hours. These three amounts would constitute the three levels of sleep deprivation, your independent variable.

To *manipulate* your independent variable, you must expose your participants to at least two levels of that variable. The specific conditions associated with each level are called the **treatments** of the experiment. Depending on the design of your experiment, the independent variable may be manipulated by exposing a different group of participants to each treatment or by exposing each participant to all the treatments in sequence. By manipulating the independent variable, you hope to show that changes in the level of the independent variable *cause* changes in the behavior being recorded.

The variable whose value you observe and measure in experimental designs is called the **dependent variable** (or *dependent measure*). If a causal relationship exists, then the value of the dependent variable depends, at least to some extent, on the level of the independent variable. (Its value also depends on other factors such as participant characteristics.) Another way to think about the dependent variable is that its value depends on the behavior of the participant, rather than being set by the experimenter.

Manipulating an independent variable can be as simple as exposing one group of participants to some treatment (e.g., distracting noises) and another group of participants to the absence of the treatment (no distracting noise). In this most basic of experimental designs, the group receiving the treatment is called the **experimental group** and the other group the **control group**. The control group is treated exactly like the experimental group except that it is not exposed to the experimental treatment. The performance of the participants in the control group provides a baseline of behavior against which the behavior of the participants in the experimental groups is compared.

Although all experiments present at least two levels of the independent variable, many do not include a no-treatment control group. A clinical study, for example, might compare a standard therapy with a new, experimental therapy of unknown effectiveness. Administering the standard therapy to the control group ensures that even the participants who do not receive the experimental treatment do not go untreated for their disorder. In both cases, the behavior of participants in the control group provides a baseline against which to compare the behavior of participants in the experimental group.

More complex experiments can be conducted using more levels of the independent variable, several independent variables, and several dependent variables. You also can choose to expose a single group, or even a single participant, to several levels of an independent variable.

**Control Over Extraneous Variables** The second characteristic of experimental research is control over extraneous variables. **Extraneous variables** are those that may affect the behavior that you wish to investigate but are not of interest for the present experiment. For example, you may be interested in determining how well a new anxiety therapy (experimental group), compared with an existing therapy (control group), affects test anxiety in anxious students. If some of your participants show up for the experiment drunk, their degree of intoxication becomes an extraneous variable. This would be especially problematic if more drunk students ended up in one group than in the other.

If allowed to vary on their own, extraneous variables can produce uncontrolled changes in the value of the dependent variable, with two rather nasty possible consequences. First, uncontrolled variability may make it difficult or impossible to detect any effects of the independent variable. (In our example, the effects of the therapy could be buried under the effects of the alcohol.) Second, uncontrolled variability may produce chance differences in behavior across the levels of the independent variable. These differences could make it appear as though the independent variable produced effects when it did not (the therapy would appear to work even though the real effect came from the alcohol). To identify clear causal relationships between your independent and dependent variables, you must control the effects of extraneous variables.

You have two ways to control these effects. The first way is simply to *hold extraneous variables constant*. If these variables do not vary over the course of your experiment, they cannot cause uncontrolled variation in your dependent variable. In the test anxiety experiment, for example, you might want to make sure that all your participants are sober (or at least intoxicated to the same degree). In fact, to the degree possible, you would want to make sure that all treatments are *exactly* alike, except for the level of the independent variable.

The second way to deal with extraneous variables is to *randomize their effects across treatments*. This technique deals with the effects of extraneous variables that cannot be held constant or, for reasons that will be explained later, should not be held constant. In an experiment assessing the effect of sleep deprivation on memory, for example, it may not be possible to ensure that all your participants have had identical amounts of sleep deprivation (some may have slept better than others the day before your experiment began) or that their recall abilities are equivalent. The idea is to distribute the effects of these differences across treatments in such a way that they tend to even out and thus cannot be mistaken for effects of the independent variable.

For statistical reasons, one of the better ways to accomplish this goal is to use random assignment of subjects to treatments. With **random assignment**, you assign participants to treatments randomly by picking their names out of a hat, for example. (In practice, one does not use names in a hat.) A table of random numbers can be used to assign subjects to treatment conditions randomly. Random assignment does not *guarantee* that the effects of extraneous variables will be distributed evenly across

treatments, but it usually works reasonably well; better yet, it allows you to use inferential statistics to evaluate the probability with which chance alone could have produced the observed differences. (We discuss the logic underlying inferential statistics in Chapter 14.) Other techniques to deal with uncontrolled extraneous variables are also available. We describe these in later chapters that cover specific design options.

However it is done, control over extraneous variables is crucial to establishing clear causal relationships between your variables. By controlling variables that might affect your dependent variable, you rule them out as possible alternative explanations for your results.

### An Example of Experimental Research: Cell Phone Use While Driving

As an illustration of experimental research, consider a follow-up study conducted by David Strayer and Frank Drews (2007), whose earlier research we summarized briefly in Chapter 1. The earlier research had shown that cell phone use seriously impairs performance in a simulated driving task. In the 2007 study, Strayer and Drews tested the hypothesis that “cell-phone conversations impair driving by inducing a form of inattention blindness in which drivers fail to see objects in their driving environment when they are talking on a cell phone” (Strayer & Drews, 2007, p. 128). Participants drove in a simulator that closely resembled the interior of a Ford Crown Victoria and offered a realistic view of a simulated road through the front and side windows. A video system monitored the driver’s eye movements. In one experiment, some participants drove while conversing on a hands-free cell phone; others drove without conversing. (Participants were randomly assigned to the conditions.) After completing the driving course, the drivers were tested for recognition of objects in the scenery they had “passed” along the way. The analysis focused on those objects on which the drivers’ eyes had fixated during the drive. Those drivers who had been conversing on the cell phone while driving recognized significantly fewer objects than those who had been driving without conversing. Based on this finding and others from the study, Strayer and Drews concluded that “these data support an inattention-blindness interpretation wherein the disruptive effects of cell-phone conversations on driving are due in large part to the diversion of attention from driving to the phone conversation.” (p. 128).

**Assessing the Strayer and Drews Experiment** Have you identified the features of the Strayer and Drews (2007) experiment that qualify it as a true experiment? If you have not done so yet, do it now before you read the next paragraphs.

A crucial element of every true experiment is the manipulation of at least one independent variable. What is the independent variable in the Strayer and Drews (2007) study? If you said that the presence or absence of a cell phone conversation while driving was the independent variable, you are correct. Note that the value of the independent variable to which a given participant was exposed (cell phone conversation or no conversation) was assigned by the experimenters; it was not chosen by the participant.

A second crucial element in an experiment is measuring a dependent variable. Can you identify the dependent variable in Strayer and Drews’ (2007) experiment? If you said that the ability to recall details about the objects on which the driver fixated

was the dependent variable, you are correct. Notice that Strayer and Drews were looking for changes in the value of the dependent variable relating to changes in the value of the independent variable.

A third crucial element of an experiment is control over extraneous variables. Were extraneous variables controlled in the Strayer and Drews (2007) experiment and, if so, how? The answer to the first part of this question is yes, and if you examine the design of the study carefully, you will see that extraneous variables were controlled using both methods described earlier. First, several extraneous variables were held constant across treatments. For example, all drivers used the same simulator and saw identical scenery along the “route.” And other than the use of a cell phone or not, both groups of participants received the same treatment. Second, the participants were assigned to their treatments randomly, not according to some behavior or characteristic of the participants. This design ensured that any remaining uncontrolled differences in the participants would tend to be distributed evenly between the two treatments. As a result, the investigators could be reasonably sure that any differences found between treatments in the values of the dependent measures were caused by the difference in treatments—that is, by the difference between holding a conversation on a cell phone while driving and not doing so.

### Strengths and Limitations of the Experimental Approach

The great strength of the experimental approach is its ability to identify and describe causal relationships. This ability is not shared by the correlational approach. Whereas the correlational approach can tell you only that changes in the value of one variable tend to accompany changes in the value of a second variable, the experimental approach can tell you whether changes in one variable (the independent variable) actually caused changes in the other (the dependent variable).

Despite its power to identify causal relationships, the experimental approach has limitations that restrict its use under certain conditions. The most serious limitation is that you cannot use the experimental method if you cannot manipulate your hypothesized causal variables. For example, studies of personality disorders must use correlational approaches to identify possible causal relationships. Exposing people to various nasty conditions in order to identify which of those conditions cause personality disorders is not ethical.

A second limitation of the experimental approach entails the tight control over extraneous factors required to clearly reveal the effects of the independent variable. Such control tends to reduce your ability to apply your findings to situations that differ from the conditions of your original experiment. A rather unpleasant trade-off exists in experimental research: As you increase the degree of control that you exert over extraneous variables (and thus your ability to establish causal relationships), you decrease your ability to assess the generality of any relationships you uncover. For example, in the Strayer and Drews (2005) experiment, extraneous variables such as simulated traffic and scenery were controlled. However, this control may limit the generality of their results because it is possible that different results would be obtained using other traffic scenarios that are, for example, more or less demanding. (We discuss the problem of generality more fully later in the chapter.)



## Experiments Versus Demonstrations

One kind of research design resembles an experiment but lacks one of the crucial features of a true experiment, an independent variable. This design, called a **demonstration**, exposes a group of subjects to one (and only one) treatment condition. Remember, a true experiment requires exposing subjects to at least two treatments. Whereas a true experiment shows the effect of manipulating an independent variable, a demonstration simply shows what happens under a specified set of conditions. To conduct a demonstration, you simply expose a single group to a particular treatment and measure the resulting behavior.

Demonstrations can be useful because they show that, under such-and-such conditions, *this* happens and not *that*. However, demonstrations are not experiments and thus do not show causal relationships. This fact is sometimes overlooked as the following example shows.

In his book *Subliminal Seduction* (1973), Wilson Bryan Key reported a study in which the participants looked at a Gilbey's Gin advertisement that allegedly had subliminal sexual messages embedded within it. The most prominent subliminal message was the word "SEX" spelled out in the bottom three ice cubes in the glass to the right of a bottle of gin (Key, 1973).

Key (1973) reported that the ad was tested "with over a thousand subjects" (the details of the study were not given). According to Key, 62% of the male and female participants reported feelings of sexual arousal in response to the ad. Key concluded that the subliminal messages led to sexual arousal. Key asserted that advertisers capitalize on these subliminal messages to get you to buy their products.

Are you convinced of the power of subliminal messages by this demonstration? If you said you are not convinced, good for you! The fact that 62% of the participants reported arousal is *not* evidence that the subliminal messages caused the arousal, no matter how many participated. All you know from this demonstration is that under the conditions tested, the advertisement evoked reports of arousal in a fair proportion of the participants. You do not learn the cause.

In fact, several plausible alternatives can be offered to the explanation that the arousal was caused by subliminal perception. For example, an advertisement for alcohol may lead participants to recall how they feel when under the influence or may conjure up images of having fun at a party. As the demonstration was reported, you cannot tell which of the potential explanations is valid. What would you have to do to fully test whether subliminal messages (such as the ones in the Gilbey's Gin ad) actually lead to sexual arousal? Give this question some thought before continuing.

To test whether subliminal messages caused the arousal, you need to add a control group and randomly assign participants to groups. Participants in this control group would see the same Gilbey's Gin ad but without the subliminal messages. If 62% of the participants in the "subliminal" group were aroused but only 10% in the control group were aroused, then you could reasonably conclude that the subliminal messages caused the arousal. A different conclusion would be drawn if 62% of the participants in *both* groups reported arousal. In this case, you would have to conclude that the subliminal messages were ineffective. The fact that the ad leads to reports of sexual arousal (as shown by the demonstration) would have to be explained by some

other factor. By the way, most of the controlled, scientific research on subliminal perception shows little or no effect of subliminal messages on behavior.

## QUESTIONS TO PONDER

---

1. What are the characteristics of experimental research?
2. What is the relationship between an independent and a dependent variable in an experiment?
3. How do extraneous variables affect your research?
4. What can be done to control extraneous variables?
5. How does a demonstration differ from a true experiment?
6. What is the value of doing a demonstration?

## INTERNAL AND EXTERNAL VALIDITY

---

Whether the general design of your study is experimental or correlational, you need to consider carefully two important but often conflicting attributes of any design: internal and external validity. In this section, we define these concepts and briefly discuss the factors that you should consider relating to internal and external validity when choosing a research design.

### Internal Validity

Much of your research will be aimed at testing the hypotheses you developed long before you collected any data. The ability of your research design to adequately test your hypotheses is known as its **internal validity** (Campbell & Stanley, 1963). Essentially, internal validity is the ability of your design to test the hypothesis that it was designed to test.

In an experiment, this means showing that variation in the independent variable, and only the independent variable, caused the observed variation in the dependent variable. In a correlational study, it means showing that changes in the value of your criterion variable relate solely to changes in the value of your predictor variable and not to changes in other, extraneous variables that may have varied along with your predictor variable.

Internal validity is threatened to the extent that extraneous variables can provide alternative explanations for the findings of a study, or as Huck and Sandler (1979) call them, *rival hypotheses*. As an example, imagine that an instructor wants to know whether a new teaching method works better than the traditional method used with students in an introductory psychology course. The instructor decides to answer this question by using the new method to teach her morning section of introductory psychology and using the traditional method to teach her afternoon section. Both sections will use the same text, cover the same material, and receive the same tests. The effectiveness of the two methods will be assessed by comparing the average

scores achieved on the test by the two sections. Now, imagine that the instructor conducts the study and finds that the section receiving the new method receives a substantially higher average grade than the section receiving the traditional method. She concludes that the new method is definitely better for teaching introductory psychology. Is she justified in drawing this conclusion?

The answer, as you probably suspected, is no. Several rival hypotheses cannot be eliminated by the study, explanations at least as credible as the instructor's view that the new method was responsible for the observed improvement in average grade. Consider the following rival hypotheses:

1. The morning students did better because they were "fresher" than the afternoon students.
2. The morning students did better because their instructor was "fresher" in the morning than in the afternoon.
3. The instructor expected the new method to work better and thus was more enthusiastic when using the new method than when using the old one.
4. Students who registered for the morning class were more motivated to do well in the course than those who registered for the afternoon class.

These rival hypotheses do not exhaust the possibilities; perhaps you can think of others. Because the study was not designed to rule out these alternatives, there is no way to know whether the observed difference between the two sections in student performance was due to the difference in teaching methods, instructor enthusiasm, alertness of the students, or other factors whose levels differed across the sections. Whenever two or more variables combine in such a way that their effects cannot be separated, a **confounding** of those variables has occurred. In the teaching study, teaching method is confounded by all those variables just listed and more. Such a study lacks internal validity.

Confounding, although always a matter of concern, does not necessarily present a serious threat to internal validity. Confounding is less problematic when the confounding variable is known to have little or no effect on the dependent or criterion variable or when its known effect can be taken into account in the analysis. For example, in the teaching study, it may be possible to eliminate concern about the difference in class meeting times by comparing classes that meet at different times but use the same teaching method. Such data may show that meeting time has only a small effect that can be ignored. If meeting time had a larger effect, you could arrange your study of teaching method so that the effect of meeting time would tend to make the new teaching method appear worse than the standard one, thus biasing the results against your hypothesis. If your results still favored the new teaching method, that outcome would have occurred despite the confounding rather than because of it. Thus, a study may include confounding and still maintain a fair degree of internal validity if the effects of the confounding variable in the situation under scrutiny are known.

This is fortunate because it is often impossible to eliminate all sources of confounding in a study. For example, the instructor in our example might have attempted to eliminate confounding by having students randomly assigned to two sections meeting simultaneously. This would certainly eliminate those sources of confounding

related to any difference in the time at which the sections met, but now it would be impossible for the instructor to teach both classes. If a second instructor is recruited to teach one of the sections using the standard method, this introduces a new source of confounding in that the two instructors may not be equivalent in a number of ways that could affect class performance. Often the best that can be done is to substitute what you believe to be less serious threats to internal validity for the more serious ones.

**Threats to Internal Validity** Confounding variables occur in both experimental and correlational designs, but they are far more likely to be a problem in the latter, in which tight control over extraneous variables is usually lacking. Campbell and Stanley (1963) identify seven general sources of confounding that may affect internal validity: history, maturation, testing, instrumentation, statistical regression, biased selection of subjects, and experimental mortality (Table 4-1).

*History* may confound studies in which multiple observations are taken over time. Specific events may occur between observations that affect the results. For example, a study of the effectiveness of an advertising campaign against drunk driving might measure the number of arrests for drunk driving immediately before and after the campaign. If the police institute a crackdown on drunk driving at the same time that the advertisements air, this event will destroy the internal validity of your study.

*Maturation* refers to the effect of age or fatigue. Performance changes observed over time due to these factors may confound those due to the variables being studied. You might, for example, assess performance on a proofreading task before and after

**TABLE 4-1 Factors Affecting Internal Validity**

FACTOR	DESCRIPTION
History	Specific events other than the treatment occur between observations
Maturation	Performance changes due to age or fatigue confound the effect of treatment
Testing	Testing prior to the treatment changes how subjects respond in posttreatment testing
Instrumentation	Unobserved changes in observer criteria or instrument calibration confound the effect of the treatment
Statistical regression	Subjects selected for treatment on the basis of their extreme scores tend to move closer to the mean on retesting
Biased selection of subjects	Groups of subjects exposed to different treatments are not equivalent prior to treatment
Experimental mortality	Differential loss of subjects from the groups of a study results in nonequivalent groups

some experimental manipulation. Decreased performance on the second proofreading assessment may be due to fatigue rather than to any effect of your manipulation.

*Testing effects* occur when a pretest sensitizes participants to what you are investigating in your study. As a consequence, they may respond differently on a posttreatment measure than if no pretest were given. For example, if you measure participants' racial attitudes and then manipulate race in an experiment on person perception, participants may respond to the treatment differently than if no such pretest of racial attitudes was given.

In *instrumentation*, confounding may be introduced by unobserved changes in criteria used by observers or in instrument calibration. If observers change what counts as "verbal aggression" when scoring behavior under two experimental conditions, any apparent difference between those conditions in verbal aggression could be due as much to the changed criterion as to any effect of the independent variable. Similarly, if an instrument used to record activity of rats in a cage becomes more (or less) sensitive over time, it becomes impossible to tell whether activity is really changing or just the ability of the instrument to detect activity.

*Statistical regression* threatens internal validity when participants have been selected based on extreme scores on some measure. When measured again, scores will tend to be closer to the average in the population. Thus, if students are targeted for a special reading program based on their unusually low reading test scores, they will tend to do better, on average, on retesting even if the reading program has no effect.

*Biased selection of subjects* threatens internal validity because subjects may differ initially in ways that affect their scores on the dependent measure. Any influence of the independent variable on scores cannot be separated from the effect of the preexisting bias. This problem typically arises when researchers use preexisting groups in their studies rather than assigning subjects to groups at random. For example, the effect of a program designed to improve worker job satisfaction might be evaluated by administering the program to workers at one factory (experimental group) and then comparing the level of job satisfaction of those workers to that of workers at another factory where the program was not given (control group). If workers given the job satisfaction program indicate more satisfaction with their jobs, is it due to the program or to preexisting differences between the two groups? There is no way to tell.

Finally, *experimental mortality* refers to the differential loss of participants from groups in a study. For example, imagine that some people drop out of a study because of frustration with the task. A group exposed to difficult conditions is more likely to lose its frustration-intolerant participants than one exposed to less difficult conditions. Any differences between the groups in performance may be due as much to the resulting difference in participants as to any difference in conditions.

**Enhancing Internal Validity** The time to be concerned with internal validity is during the design phase of your study. During this phase, you should carefully plan which variables will be manipulated or observed and recorded, identify any plausible rival hypotheses not eliminated in your initial design, and redesign so as to eliminate those that seriously threaten internal validity. Discovering problems with internal

validity after you have run your study is too late. A poorly designed study cannot be fixed later on.

### External Validity

A study has **external validity** to the degree that its results can be extended (generalized) beyond the limited research setting and sample in which they were obtained. A common complaint about research using white rats or college students and conducted under the artificial conditions of the laboratory is that it may tell us little about how white rats and college sophomores (let alone animals or people in general) behave under the conditions imposed on them in the much richer arena of the real world.

The idea seems to be that all studies *should* be conducted in such a way that the findings can be generalized immediately to real-world situations and to larger populations. However, as Mook (1983) notes, it is a fallacy to assume “that the purpose of collecting data in the laboratory is to predict real-life behavior in the real world” (p. 381). Mook points out that much of the research conducted in the laboratory is designed to determine one of the following:

1. Whether something *can* happen, rather than whether it typically *does* happen
2. Whether something we specify *ought* to happen (according to some hypothesis) under specific conditions in the lab *does* happen there under those conditions
3. What happens under conditions not encountered in the real world

In each of these cases, the objective is to gain insight into the underlying mechanisms of behavior rather than to discover relationships that apply under normal conditions in the real world. It is this understanding that generalizes to everyday life, not the specific findings themselves.

**Threats to External Validity** In Chapter 1, we distinguished between basic research, which is aimed at developing a better understanding of the underlying mechanisms of behavior, and applied research, which is aimed at developing information that can be directly applied to solve real-world problems. The question of external validity may be less relevant in basic research settings that seek theoretical reasons to determine what will happen under conditions not usually found in natural settings or that examine fundamental processes expected to operate under a wide variety of conditions. The degree of external validity of a study becomes more relevant when the findings are expected to be applied directly to real-world settings. In such studies, external validity is affected by several factors. Using highly controlled laboratory settings (as opposed to naturalistic settings) is one such factor. Data obtained from a tightly controlled laboratory may not generalize to more naturalistic situations in which behavior occurs. Other factors that affect external validity, as discussed by Campbell and Stanley (1963), are listed and briefly described in Table 4-2. Many of these threats to external validity are discussed in later chapters, along with the appropriate research design.

**TABLE 4-2 Factors Affecting External Validity**

FACTOR	DESCRIPTION
Reactive testing	Occurs when a pretest affects participants' reaction to an experimental variable, making those participants' responses unrepresentative of the general population
Interactions between participant selection biases and the independent variable	Effects observed may apply only to the participants included in the study, especially if they are unique to a group (such as college sophomores rather than a cross section of adults)
Reactive effects of experimental arrangements	Refers to the effects of highly artificial experimental situations used in some research and the participant's knowledge that he or she is a research participant
Multiple treatment interference	Occurs when participants are exposed to multiple experimental treatments in which exposure to early treatments affects responses to later treatments

### Internal Versus External Validity

Although you should strive to achieve a high degree of both internal and external validity in your research, in practice you will find that the steps you take to increase one type of validity tend to decrease the other. For example, a tightly controlled laboratory experiment affords you a relatively high degree of internal validity. Your findings, however, may not generalize to other samples and situations; thus, external validity may be reduced. Often the best that you can do is reach a compromise on the relative amounts of internal and external validity in your research.

Whether internal or external validity is more important depends on your reasons for conducting the research. If you are most interested in testing a theoretical position (as is often the case in basic research), you might be more concerned with internal than external validity and hence conduct a tightly controlled laboratory experiment. However, if you are more concerned with applying your results to a real-world problem (as in applied research), you might take steps to increase the external validity while attempting to maintain a reasonable degree of internal validity. These issues need to be considered at the time when you design your study.

As just mentioned, the setting in which you conduct your research strongly influences the internal and external validity of your results. The kinds of setting available and the issues that you should consider when choosing a research setting are the topics that we take up next.

## QUESTIONS TO PONDER

---

1. What is internal validity, and why is it important?
2. What factors threaten internal validity?
3. How do confounding variables threaten internal validity, and how can they be avoided?
4. What is external validity, and when is it important to have high levels of external validity?
5. How do internal and external validity relate to one another?

## RESEARCH SETTINGS

---

In addition to deciding on the design of your research, you also must decide on the setting in which you conduct your research. Your choice of setting is affected by the potential costs of the setting, its convenience, ethical considerations, and the research question that you are addressing.

The two research settings open for psychological research are the laboratory and the field. For this discussion, the term *laboratory* is used in a broad sense. A laboratory is any research setting that is artificial relative to the setting in which the behavior naturally occurs. This definition is not limited to a special room with special equipment for research. A laboratory can be a formal lab, but it also can be a classroom, a room in the library, or a room in the student union building. In contrast, the *field* is the setting in which the behavior under study naturally occurs.

Your decision concerning the setting for your research is an important one, so you must be familiar with the relative advantages and disadvantages of each.

### The Laboratory Setting

If you choose to conduct your research in a *laboratory setting*, you gain important control over the variables that could affect your results. The degree of control depends on the nature of the laboratory setting. For example, if you are interested in animal learning, you can structure the setting to eliminate virtually all extraneous variables that could affect the course of learning. This is what Ivan Pavlov did in his investigations of classical conditioning. Pavlov exposed dogs to his experimental conditions while the dogs stood in a sound-shielded room. The shielded room permitted Pavlov to investigate the impact of the experimental stimuli free from any interfering sounds. Like Pavlov, you can control important variables within the laboratory that could affect the outcome of your research.

Complete control over extraneous variables may not be possible in all laboratory settings. For example, if you were administering your study to a large group of students in a psychology class, you could not control all the variables as well as you might wish (students may arrive late, or disruptions may occur in the hallway). For the most part, the laboratory affords more control over the research situation than does the field.



**Simulation: Re-creating the World in the Laboratory** When you choose the laboratory as your research setting, you gain control over extraneous variables that could affect the value of your dependent variable. However, you make a trade-off when choosing the laboratory. Although you gain better control over variables, your results may lose some generality (the ability to apply your results beyond your specific laboratory conditions). If you are concerned with the ability to generalize your results, as well as with controlling extraneous variables, consider using a **simulation**. In a simulation, you attempt to re-create (as closely as possible) a real-world situation in the laboratory. Carefully designed and executed simulation may increase the generality of results. Because this strategy has been used with increasing frequency lately, a detailed discussion is in order.

**Why Simulate?** You may decide for a variety of reasons to simulate rather than conduct research in the real world. You may choose simulation because the behavior of interest could not be studied ethically in the real world. For example, Chapter 1 mentioned factors that control panic behavior. Re-creating a panic situation in order to study the ensuing behavior is unethical. If you were interested in studying how juries reach a decision, you could not eavesdrop on real juries. However, you could conduct a jury simulation study and analyze the deliberations of the simulated juries.

Often researchers choose to simulate for practical reasons. A simulation may be used because studying a behavior under its naturally occurring conditions is expensive and time consuming. By simulating in the laboratory, the researcher also gains the advantage of retaining control over variables while studying the behavior under relatively realistic conditions.

**Designing a Simulation** For a simulation to improve the generality of laboratory-based research, it must be properly designed. Observe the actual situation and study it carefully (Winkel & Sasanoff, 1970). Identify the crucial elements and then try to reproduce them in the laboratory. The more realistic the simulation, the greater are the chances that the results will be applicable to the simulated real-world phenomenon. As an example, suppose you were interested in studying the interpersonal relationships and dynamics that evolve in prisons. It might be difficult to conduct your study in an actual prison, so you might consider a simulation. In fact, Haney, Banks, and Zimbardo (1973) did just that.

In their now-famous Stanford prison study, Haney et al. (1973) constructed a prison in the basement of the psychology building at Stanford University. Participants in the study were randomly assigned to be either prisoners or prison guards. Those participants assigned to be prisoners were “arrested” by the police, fingerprinted, and incarcerated in the simulated prison. Treatment of the prisoner-participants was like that of actual prisoners: They were issued numbers and drab uniforms and were assigned to cells. Prison guards were issued uniforms, badges, and nightsticks. Their instructions were to maintain order within the simulated prison.

The behavior of the participants within the simulated prison was observed by a team of social psychologists. Behavior within the simulated prison was similar to (though less extreme than) behavior in a real prison. Guards developed rigid and

sometimes demeaning rules, and prisoners banded together in a hunger strike. In fact, the simulation was so real for the participants that the experiment had to be discontinued after only a few days.

**Realism** Most researchers would agree that a simulation should be as realistic as possible (as was the case in the Stanford prison study). The physical reality created in the Stanford prison study probably helped participants become immersed in their roles. However, a simulation may not have to be highly realistic to adequately test a hypothesis. For example, many jury simulation studies do not re-create the physical setting of a courtroom. However, many of these studies are highly involving and compelling for the participants.

The importance of the “realism” of a simulation depends in part on the definition of *realism* that you adopt. Aronson and Carlsmith (1968) distinguish between two types of realism: mundane and experimental. The term *mundane realism* refers to the degree to which a simulation mirrors the real-world event. In contrast, *experimental realism* refers to the degree to which the simulation psychologically involves the participants in the experiment.

Simulation is an important issue in the area of social psychology and law. Many researchers have used simulation methods to study issues such as plea bargaining and jury decision making. A simulation in which a courtroom is realistically reconstructed in the laboratory could have high mundane realism. However, such high levels of mundane realism do not guarantee that the results of the study will be any more valid than those of the same study conducted in a more ordinary laboratory setting. Experimental realism is an important factor to be considered. An involving task in a laboratory with low mundane realism may produce more general results than a less involving task in a laboratory with high mundane realism.

A good illustration of the importance of experimental realism comes from a study by Wilson and Donnerstein (1977). These researchers report that a crucial factor in the applicability of simulated jury research findings is whether or not the participant believes that his or her decision will have real consequences. As an independent variable, Wilson and Donnerstein varied whether or not participants believed that their decisions would have consequences. They found that when participants believed that their judgments had consequences, the defendant’s character (a variable previously shown in other research to be an important factor in the decision process) was no longer important.

Leading the participant to believe that his or her decision has consequences beyond the advancement of science increases experimental realism and thus increases the generality of the results. You may be able to increase the generality of your results when designing simulation studies by taking steps to increase not only mundane realism but also experimental realism.

To summarize, the laboratory approach to research has the advantage of allowing you to control variables and thus to isolate the effects of the variables under study. However, in gaining such control over variables, you lose a degree of generality of results. Using simulations that are high in experimental realism may improve the ability to generalize laboratory results in the real world.

## The Field Setting

*Field research* is research conducted outside the laboratory in the participants' natural environment (the "field"). In this section, we briefly discuss conducting experiments in the field. However, most field research employs nonexperimental (correlational) methods such as naturalistic observation or survey designs. (We discuss these nonexperimental methods in Chapters 8 and 9.)

**The Field Experiment** A *field experiment* is an experiment conducted in the participant's natural environment. In a field experiment (as in a laboratory experiment), you manipulate independent variables and measure a dependent variable. You decide which variables to manipulate, how to manipulate them, and when to manipulate them. Essentially, the field experiment has all the qualities of the laboratory experiment except that the research is conducted in the real world rather than in the artificial laboratory setting.

As an example, consider an experiment conducted by Ute Gabriel and Rainer Banse (2006) to investigate whether gays and lesbians are the target of discrimination. Their measure of discrimination was whether gays and lesbians were helped less than heterosexuals. Residents of Berlin, Germany, were called between 6:00 P.M. and 9:00 P.M. over a 4-week period by a male or female researcher. The sex of the caller was communicated to participants by having the male researcher call himself Michael and the female researcher call herself Anna. Once a participant was on the telephone, the researcher asked the participant if the researcher's romantic partner was at home. Sexual orientation of the caller (researcher) was manipulated by having the caller ask for a same-sex (e.g., Michael asks for Peter) or opposite-sex partner (e.g., Anna asks for Peter). When the participant indicated that the caller had reached the wrong number, the researcher went on to explain that his or her car had broken down and that he or she did not want the romantic partner to worry. The participant was told further that the caller had no more money for another call and asked the participant to call his or her partner so that he or she would not worry. At this point, the caller gave the participant a number to call. The dependent variable was the number of participants in each experimental condition who made the call.

Gabriel and Banse (2006) found that homosexual callers were significantly less likely to receive help (67%) than heterosexual callers (83.5%). This difference was found for both male and female callers. They also found that male participants were significantly less likely to help homosexual callers than were female participants. Interestingly, Gabriel and Banse also report that male and female participants discriminated against lesbian callers at about the same rate. However, male participants discriminated against gay callers significantly more than female participants.

This field experiment has all the elements of a true experiment. Independent variables were manipulated (sex of caller and sexual orientation of caller) and a dependent variable was measured (whether the participant called the number provided by the caller). Hence, causal inferences about helping behavior can be made from the observations.

**Advantages and Disadvantages of the Field Experiment** As with the laboratory experiment, the field experiment has its advantages and disadvantages. Because the research is conducted in the real world, one important advantage is that the results can be easily generalized to the real world (i.e., high external validity). An important disadvantage is that you have little control over potential confounding variables (i.e., low internal validity). In the Gabriel and Banse (2006) field experiment, for example, the researchers could not control who would answer the telephone when the researcher called. Nor could they control how many others were present with the participant when called and what participants were doing when the call came in. Each of these variables could affect the reaction of a person asked to make a call for someone else. These extraneous variables can obscure or distort the effects of the independent variables manipulated in field experiments.

### A Look Ahead

At this point, you have been introduced to the broad issues that you should consider when choosing a research design, the basic design options available to you, and the strengths and weaknesses of each choice. Before you are ready to conduct your first study, you also will need to know how to measure your variables; what methods of observation are available; how to conduct systematic, reliable, and objective observations; how to choose participants and deal with them ethically; how to minimize participant and experimenter biases; and many other details concerning specific research designs. In the next chapter, we consider how to go about making systematic, scientifically valid observations.

### QUESTIONS TO PONDER

---

1. What is a simulation, and why would you use one?
2. How does the realism of a simulation relate to the validity of the results obtained from a simulation?
3. What are the defining features of laboratory and field research?
4. What are the relative advantages and disadvantages of laboratory and field research?

### SUMMARY

---

Some of the most important decisions that you will make about your research concern its basic design and the setting in which it will be conducted. Research designs serve one or both of two major functions: (1) exploratory data collection and analysis (to identify new phenomena and relationships) and (2) hypothesis testing (to check the adequacy of proposed explanations). In the latter case, it is particularly important to distinguish causal from correlational relationships between variables. The relationship is causal if one variable directly influences the other.

The relationship is correlational if the two variables simply change values together (covary) and may or may not directly influence one another.

Two basic designs are available for determining relationships between variables: correlational designs and experimental designs. Correlational research involves collecting data on two or more variables across subjects or time periods. The states of the variables are simply observed or measured “as is” and not manipulated. Participants enter a correlational study already “assigned” to values of the variables of interest by nature or circumstances. Correlational designs can establish the existence of relationships between the observed variables and determine the direction of the relationships. However, two problems prevent such designs from determining whether the relationships are causal. The third-variable problem arises because of the possibility that a third, unmeasured variable influences both observed variables in such a way as to produce the correlation between them. The directionality problem arises because, even if two variables are causally related, correlational designs cannot determine in which direction the causal arrow points.

Despite its limitations, correlational research is useful on several accounts. It provides a good method for identifying potential causal relationships during the early stages of a research project, can be used to identify relationships when the variables of interest cannot or should not be manipulated, and can show how variables relate to one another in the real world outside the laboratory. Such relationships can be used to make predictions even when the reasons for the correlation are unknown. A variable in a correlational relationship that is used to make predictions is termed a *predictor variable*, and a variable whose value is being predicted is termed a *criterion variable*.

Experimental designs provide strong control over variables and allow you to establish whether variables are causally related. The defining characteristics of experimental research are (1) manipulation of an independent variable and (2) control over extraneous variables. Independent variables are manipulated by exposing subjects to different values or levels and then assessing differences in the participants’ behavior across the levels. The observed behavior constitutes the dependent variable of the study. Extraneous variables are controlled by holding them constant, if possible, or by randomizing their effects across the levels of the independent variable.

The simplest experimental designs involve two groups of participants. The experimental group receives the experimental treatment; the control group is treated identically except that it does not receive the treatment. More complex designs may include more levels of the independent variable, more independent variables, or more dependent variables.

Although experiments can identify causal relationships, in some situations they cannot or should not be used. Variables may be impossible to manipulate, or it may be unethical to do so. In addition, tight control over extraneous variables may limit the generality of the results.

A demonstration is a type of nonexperimental design that resembles an experiment but lacks manipulation of an independent variable. It is useful for showing what sorts of behaviors occur under specific conditions, but it cannot identify relationships among variables.

Two important characteristics of any design are its internal and external validity. Internal validity is the ability of a design to test what it was intended to test. Results

from designs low in internal validity are likely to be unreliable. A serious threat to internal validity comes from confounding. Confounding exists in a design when two variables are linked in such a way that the effects of one cannot be separated from the effects of the other. External validity is the ability of a design to produce results that apply beyond the sample and situation within which the data were collected. Results from designs low in external validity have little generality when applied directly to real-world situations. However, not all research is designed for such application; non-applied studies need not possess high external validity.

After deciding on a research design, you must then decide on a setting for your research. You can conduct your research in the laboratory or in the field. The laboratory setting affords you almost total control over your variables. You can tightly control extraneous variables that might confound your results. Laboratory studies, however, tend to have a degree of artificiality. You cannot be sure that the results you obtain in the laboratory apply to real-world behavior. Simulation is a technique in which you seek to re-create the setting in which the behavior naturally occurs. The success of your simulation depends on its realism, which is of two types. Mundane realism is the degree to which your simulation re-creates a real-world environment. Experimental realism concerns how involved in your study your participants become. High levels of mundane realism do not guarantee a valid simulation. Experimental realism is often more important.

Field research is conducted in your participants' natural environment. Although this setting allows you to generalize your results to the real world, you lose control over extraneous variables. Field experiments therefore tend to have high external validity but relatively low internal validity.

## KEY TERMS

---

causal relationship

correlational relationship

correlational research

third-variable problem

directionality problem

experimental research

independent variable

treatments

dependent variable

experimental group

control group

extraneous variable

random assignment

demonstration

internal validity

confounding

external validity

simulation

# Making Systematic Observations

The everyday observations that we make (the weather is hot and humid today; Martha is unusually grouchy; I'm feeling grouchy, too) are generally unsystematic, informal, and made haphazardly, without a plan. In contrast, scientific observations are systematic: What will be observed, how the observations will be made, and when the observations will be made are all carefully planned in advance of the actual observation.

Information recorded in this systematic way becomes the data of your study. Your conclusions come from these data, so it is important that you understand how your choice of variables to observe, methods of measurement, and conditions of observation affect the conclusions you can legitimately draw. This chapter provides the information you need to make these choices intelligently.

## DECIDING WHAT TO OBSERVE

In Chapters 3 and 4, we discussed how to obtain and develop a research idea and how to select a general strategy to attack the questions your research idea raises. After you select a specific question to investigate, you must decide exactly what to observe. Most research situations offer many ways to address a single question.

As one example, assume that you want to study the relationship between weather and mood. Your general research question involves how the weather relates to a person's mood. You must decide what specific observations to make. First, you must specify what you mean by *weather*. *Weather* can be defined in terms of a number of specific variables, such as barometric pressure, air temperature and humidity, amount of sunlight, and perhaps the type and amount of precipitation. You may want to measure and record all these variables, or you may want to define *weather* in terms of some combination of these variables. For example, you could dichotomize weather into

## CHAPTER OUTLINE

### Deciding What to Observe

#### Choosing Specific Variables for Your Study

- Research Tradition

- Theory

- Availability of New Techniques

- Availability of Equipment

#### Choosing Your Measures

- Reliability of a Measure

- Accuracy of a Measure

- Validity of a Measure

- Acceptance as an Established Measure

- Scale of Measurement

- Variables and Scales of Measurement

- Choosing a Scale of Measurement

- Adequacy of a Dependent Measure

- Tailoring Your Measures to Your Research Participants

- Types of Dependent Variables and How to Use Them

#### Choosing When to Observe

#### The Reactive Nature of Psychological Measurement

- Reactivity in Research with Human Participants

- Demand Characteristics

- Other Influences

- The Role of the Experimenter

- Reactivity in Research with Animal Subjects

#### Automating Your Experiments

#### Detecting and Correcting Problems

- Conducting a Pilot Study

#### Adding Manipulation Checks

#### Summary

#### Key Terms

two general categories: gloomy (cloudy or foggy, humid, low barometric pressure) and zesty (sunny, dry, high barometric pressure).

You also must decide how to index the moods of your participants. Again, a number of possibilities exist. You may choose to have participants rate their own moods, perhaps by using the Mood Adjective Check List (Nowlis & Green, 1957, cited in Walster, Walster, & Berscheid, 1978), or you may decide to gauge the moods of your participants through observation of mood-related behaviors. In this example, you have translated your general research idea into action by selecting particular observations to make. Note that the same general variables (weather, mood) can be defined and measured in a number of ways. As discussed in Chapter 3, the specific way that you choose to measure a variable becomes the *operational definition* of that variable within the context of your study. How you choose to operationalize a variable, and thus to observe and measure it, affects how you will later analyze your data and determines what conclusions you can draw from that analysis. So you should carefully consider what variables to observe and manipulate and how to operationally define them.

## CHOOSING SPECIFIC VARIABLES FOR YOUR STUDY

---

Assuming that you have decided on a general research topic, a number of factors may influence your choice of specific variables to observe and manipulate. Some of them are research tradition, theory, availability of new techniques, and availability of equipment.

### Research Tradition

If your topic follows up on previous research in a particular area, the variables that you choose to observe may be the same as those previously studied. In particular, you may choose to study the same dependent variables while manipulating new independent variables. For example, research on operant conditioning typically focuses on how various factors affect the rate of lever pressing (in rats) or key pecking (in pigeons). In experiments on cognitive processing, reaction times are frequently recorded to determine how long a hypothesized process requires to complete. Using these traditional measures allows you to compare the results of different manipulations across experiments.

### Theory

Your decision about what to observe may depend on a particular theoretical point of view. For example, you may choose to observe behaviors that are seen as important from a certain theoretical perspective. If these behaviors (or other variables) have been used in previous research, you probably should use the measures already developed for them. However, the theory may suggest looking at behaviors not previously observed, in which case you may need to develop your own measures.



### Availability of New Techniques

Sometimes a variable cannot be investigated because there is no suitable way to measure it. In this case, the development of new techniques may open the way to observation and experimentation. You may want to use the new measure simply to explore its potential for answering your research question. As an example, consider the development of positron emission tomography (PET), a technique allowing researchers to visualize the level of activity of parts of a person's brain. A scanner picks up positrons (positively charged electrical particles) emitted by radioactively labeled glucose, which is being absorbed by neurons of the cerebral cortex to fuel their metabolic activity. More active neurons absorb more glucose and therefore emit more positrons. A computer translates the rates of positron emission in various regions of the cortex into a color-coded image of the cortex on the computer's display screen. By keeping track of changes in the colors, an observer can determine the ongoing pattern of neural activity.

This technology has enabled researchers to observe which parts of the cortex are most active during a variety of cognitive tasks. For example, using PET technology, Hakan Fischer, Jesper Anderson, Thoms Furmark, Gustav Wik, and Mats Fredrikson (2002) found increased metabolic activity in the right medial gyrus of the prefrontal cortex when an individual was presented with a fear-inducing stimulus. No such activity was found when an individual was presented with a nonfear control stimulus. Thus, using PET scan technology, Fischer et al. could confirm the role of the prefrontal cortex in mediating fear responses.

### Availability of Equipment

You are always tempted to adopt measures for which you already are equipped. For example, if you have invested in an operant chamber equipped with a lever and feeder, you may find it easier to continue your studies of operant conditioning by using this equipment rather than starting from scratch. Perhaps this equipment makes it trivially easy to collect data on response frequency (number of lever presses per minute) but does not readily yield information about response duration (amount of time the lever is depressed) or response force (amount of pressure exerted on the lever). You may decide that measuring response frequency will be adequate to answer your research question, particularly if previous research has successfully used this measure.

If the chosen measures provide reasonable answers to your research questions, this decision is not wrong. Problems arise when the measure really is not appropriate or adequate for the question being investigated but is chosen anyway on the basis of mere convenience. If you have chosen a particular measure simply because it is readily available or convenient, you should ask yourself whether it really is the best measure for your question.

The decision of how to observe the behavior and other variables of your study requires that you select appropriate measures of these variables. In the next section, we examine some issues that you need to consider when choosing measures of your variables.

## QUESTION TO PONDER

---

What factors should you consider when deciding what to observe in a study?

## CHOOSING YOUR MEASURES

---

Whether your research design is experimental or correlational, your study will involve measuring the values of those variables included in the design. Yet there are many ways in which a given variable can be measured, and some may prove better for your purposes than others. In this section, we describe several important characteristics of a measure that you should consider before adopting it for your study, including its reliability, its accuracy, its validity, and the level of measurement it represents. We then discuss two additional factors that affect the adequacy of a dependent measure: its sensitivity and its susceptibility to range effects. Next, we take up the problem of tailoring your measures to your research participants. Measures must be adapted to the special situations posed, for example, by the testing of young children. Finally, we identify and describe several types of behavioral measure commonly used in psychological research.

### Reliability of a Measure

The **reliability** of a measure concerns its ability to produce similar results when repeated measurements are made under identical conditions. Imagine weighing yourself several times in quick succession using an ordinary bathroom scale. You expect to see the same body weight appear on the scale each time, but if the scale is cheap or worn, the numbers may vary by 1 or 2 pounds, or even worse. The more variability that you observe, the less reliable is the measure. Procedures used to assess reliability differ depending on the type of measure, as discussed next.

**Reliability of a Physical Measure** The reliability of measures of physical variables such as height and weight are assessed by repeatedly measuring a fixed quantity of the variable and using the observed variation in measured value to derive the *precision* of the measure, which represents the range of variation to be expected on repeated measurement. For example, the precision of weighings produced by a given bathroom scale might be reported as  $\pm 1.2$  pounds. A more precise measurement has a smaller range of variation.

**Reliability of Population Estimates** For measures of opinion, attitude, and similar psychological variables, in which the problem is to estimate the average value of the variable in a given population based on a sample drawn from that population, the precision of the estimate (its likely variation from sample to sample) is called the *margin of error*. The results of a poll of registered voters asking whether the voter favors or opposes stronger legislation on gun control might be reported as “41% favor stronger legislation, 54% are against it, and 5% are unsure, with a margin of error of  $\pm 3\%$ .”

**Reliability of Judgments or Ratings by Multiple Observers** When the measure being made consists of judgments or ratings of multiple observers, you can establish the degree of agreement among observers by using a statistical measure of *interrater reliability*. (We describe ways to assess interrater reliability in Chapter 6.)

**Reliability of Psychological Tests or Measures** Assessing the reliability of measures of psychological variables such as intelligence, introversion/extraversion, anxiety level, mood, and so on poses a special difficulty in that these variables tend to change naturally over time. By the time that you repeat a measurement of mood or anxiety level, for example, the underlying quantity being measured in the individual may have changed. If so, the measure will appear to be unreliable even though the changes in measured value reflect real changes in the variable. In addition, for various reasons, it is often not possible to administer psychological assessment devices to the same individuals a sufficient number of times to determine the reliability of the measure. Thus, an alternative strategy is needed for assessing the reliability of these measures.

The basic strategy for assessing the reliability of psychological measures is to administer the assessment twice to a large group of individuals and then determine the *correlation* (Pearson  $r$ ) between the scores on the first and second administrations. The higher the correlation, the greater the reliability. A test is considered to have high reliability if  $r$  is .95 or higher. (See Chapter 13 for a discussion of the Pearson  $r$  statistic.) You can choose among several methods for assessing the reliability of a psychological test, each with a different set of advantages and drawbacks. These include the test–retest, parallel-forms, and split-half reliability assessments.

**Test–retest reliability** involves administering the same test twice, separated by a relatively long interval of time. Because the same test is used on each occasion, changes in scores on the test cannot be due to such factors as different wording of the questions or nonequivalent items. By the same token, however, participants may respond in the same way on repeated administration simply because they recall how they responded on first administration. If so, the test will appear to be more reliable than it actually is. Furthermore, participants may change between administrations of the test, leading to an artificially low reliability figure. For these reasons, the test–retest method is best for assessing stable characteristics of individuals such as intelligence. The variable being assessed by the test is unlikely to change much between administrations, and administrations can be spaced far enough apart that participants are unlikely to remember much about their previous responses to the test.

The problem of remembering previous responses can be avoided by assessing a **parallel-forms reliability** (or *alternate-forms reliability*). This is the same as test–retest reliability except that the form of the test used on first administration is replaced on second administration by a parallel form. A parallel form contains items supposedly “equivalent” to those found in the original form. These assess the same knowledge, skills, and so on but use somewhat different questions or problems, which eliminates the possibility that on second administration the person could simply recall his or her answer on the previous occasion. However, if the items of the parallel form are not really equivalent, differences in test performance due to this nonequivalence may reduce the apparent reliability of the test. In addition, the parallel-forms method still

suffers from the possibility that the quantity measured may have changed since first administration, thus making the test appear less reliable than it really is.

You can avoid the problem caused by changes between administrations in the quantity being measured by choosing the **split-half reliability** method. Here, the two parallel forms of the test are intermingled in a single test and administered together in one sitting. The responses to the two forms are then separated and scored individually. Because both forms are administered simultaneously, the quantity being measured has no time to change. However, the need to use alternate forms in the two halves introduces the same problem found in the parallel-forms method, that of ensuring that the two forms are in fact equivalent.

These methods for assessing the reliability of a psychological test apply equally well to assessing the reliability of a questionnaire designed for distribution in a survey. (For a discussion of these methods in the context of survey design, see Chapter 9.)

### Accuracy of a Measure

The term **accuracy** describes a measure that produces results that agree with a known standard. For example, a bathroom scale is accurate if it indicates 50 pounds when a standard 50-pound weight is placed on it, 100 pounds when a standard 100-pound weight is placed on it, and so on. A thermometer calibrated in degrees Celsius (C) is accurate if it reads 0 degrees when tested in a slurry of ice and 100 degrees when placed in boiling water (both tested under sea-level air pressure). A counter is accurate if the number of events counted equals the actual number of events that occurred.

Determining accuracy is hampered by lack of precision. Your measurement may not agree with the known standard each time that you make it. However, the measurement may still be accurate in the sense that the value observed agrees with the standard *on average*. Thus, you can determine accuracy by measuring the standard a large number of times and computing the average; the measure is accurate if the average value equals the value of the standard. Any difference between this average value and the standard value is termed *bias*. Bias can be overcome either by adjusting the measuring instrument to eliminate it or, if this is not possible, by mathematically removing the bias from the measured value.

Although a somewhat unreliable measure may be accurate *on average*, any *single* measurement in such a case will tend to deviate from the actual value by some amount. When a value is reported as being, for example, “accurate to within  $\pm 0.1$  centimeter” (cm), this means that, in general, measured values will tend to be within 0.1 cm of the true value. Thus, the precision of the measure limits the accuracy (probable closeness to the true value) of a single measurement. However, the converse is not true. A measurement can be precise (repeatable within narrow limits) and yet wildly inaccurate. For example, a thermometer whose glass has slipped with respect to the scale behind it may yield the same value in ice water within  $\pm 0.1$  °C, yet give an average value of 23 degrees instead of the correct 0 degrees.

In psychological measurement, standards are rare and, therefore, the accuracy of a measure cannot be assessed. This does not mean that you should ignore accuracy issues altogether. For example, no standard introvert exists against which to assess the accuracy of a measure of introversion/extraversion (a personality variable). In such

cases, test scores may be “standardized” by statistical methods to have a specified average value in a given population and a specified amount of variability. You can find an extensive discussion of these methods and other issues related to psychological testing in Cohen and Swerdlik (2010).

### Validity of a Measure

In the previous chapter, we introduced the concepts of internal and external validity, which are attributes of a research design. In this section, we discuss other forms of validity that apply to measures. The **validity** of a *measure* is the extent to which it measures what you intend it to measure. Imagine, for example, that you decided you could “measure” a person’s general intelligence by placing a tape measure around that person’s skull at the level of the forehead, on the theory that larger skulls house larger brains and that larger brains produce higher intelligence. Most of us would agree that the tape measure is a valid measure of *length*, but used in this way, is it a valid measure of intelligence? This question was actually investigated. Near the end of the 19th century, the so-called science of phrenology enjoyed a brief popularity. Phrenologists believed that by carefully measuring the cranium of a person, they could learn something about that person’s personality, aptitudes, and, yes, general intelligence. They were wrong. For one thing, over the normal range of variation (excluding pathological cases such as microcephaly) the correlations between brain size and intelligence are very small. In fact, the largest brain on record belonged to a mildly retarded person, and several of the leading thinkers of the day turned out to have disappointingly small brains. Thus, measures of brain size turned out not to be the most valid indicator of intelligence (Fancher, 1979).

You should be concerned about the validity of any measure, but in psychology the topic comes up most often when discussing tests designed to measure psychological attributes. In this context, several types of validity have been defined, each requiring a somewhat different operation to establish. Here we briefly discuss three: face validity, content validity, and criterion-related validity. (For more information on test validity, see Chapter 9.)

**Face validity** describes how well a measurement instrument (e.g., a test of intelligence) appears to measure (judging by its appearance) what it was designed to measure. For example, a test of mathematical ability would have face validity if it contained math problems. Face validity is a weak form of validity in that an instrument may lack face validity and yet, by other criteria, measure what it is intended to measure. Nevertheless, having good face validity may be important. If those who take the test do not perceive the test as valid, they may develop a negative attitude about its usefulness (Cohen & Swerdlik, 2010).

**Content validity** has to do with how adequately the content of a test samples the knowledge, skills, or behaviors that the test is intended to measure. For example, a final exam for a course would have content validity if it adequately sampled the material taught in the course. An employment test would have content validity if it adequately sampled from the larger set of job-related skills. Finally, a test designed to measure “assertiveness” would have content validity to the extent that it adequately sampled from the population of all behaviors that would be judged as “assertive” (Cohen & Swerdlik, 2010).

**Criterion-related validity** reflects how adequately a test score can be used to infer an individual's value on some "criterion" measure. To determine the test's criterion-related validity, you compare the values inferred from the test to the criterion values actually observed. Criterion-related validity includes two subtypes. You assess **concurrent validity** if the scores on your test and the criterion are collected at about the same time. For example, you might establish the concurrent validity of a new, 10-minute, paper-and-pencil test of intelligence by administering it and the Stanford–Binet (an established test of intelligence) at about the same time and demonstrating that the scores on the two tests correlated strongly. You assess **predictive validity** by comparing the scores on your test with the value of a criterion measure observed at a later time. A high correlation between these measures indicates good predictive validity. Predictive validity indicates the ability of a test to predict some future behavior. For example, the Scholastic Assessment Test (SAT), given in high school, does a good job of predicting future college performance (as shown by its high correlation with the latter) and thus has predictive validity.

Finally, **construct validity** applies when a test is designed to measure a "construct," which is a variable, not directly observable, that has been developed to explain behavior on the basis of some theory. Examples of constructs include such variables as "intelligence," "self-esteem," and "achievement motivation." To demonstrate the construct validity of a measure, you must demonstrate that those who score high or low on the measure behave as predicted by the theory. For example, those who receive low (high) scores on an intelligence test should behave the way people of low (high) intelligence would be expected to behave, as predicted by the theory of intelligence on which the construct was based.

Just as a measure can be reliable but inaccurate, it also can be reliable but invalid. The phrenologists whom we discussed earlier developed large calipers and other precision instruments to make the task of measurement reliable. By using these instruments properly, they were able to collect highly reliable measurements of cranial shapes and sizes. Unfortunately, the phrenologists chose to interpret these measurements as indicators of the magnitudes of various mental characteristics such as memory, personality, intelligence, and criminality. Of course, cranial size and shape actually provide no such information. Despite being highly reliable, the phrenologists' measures were not valid indicators of mental characteristics.

Although a measure can be reliable but invalid, the converse is not true. If a measure is unreliable, it is not a valid gauge of anything except the amount of random error in the measuring instrument.

### Acceptance as an Established Measure

In our weather and mood example, the Mood Adjective Check List was one possible measure of participants' moods. This established measure has been used in previous research. Using established measures is advantageous because the reliability and the validity of the measure are known.

Although you do not have to spend precious time validating an established measure, it may not be suitable for addressing your research questions. A case in which the established measure was *not* appropriate comes from the literature on jury decision

making. Early research on the factors that affect jury decision making required participants to sentence a defendant (e.g., see Landy & Aronson, 1969), and several subsequent studies also used this measure. Because jurors are not empowered to sentence a defendant (except in death penalty cases and a few other cases in some jurisdictions), the established measure lacked realism. Later research attempted to correct this problem by having participants evaluate the guilt of the defendant either on rating scales or as a dichotomous (two-value) guilty/not guilty verdict.

An alternative to using established measures is to develop your own. This alternative has the advantage of freeing you from previous dogma and theory. In fact, a successful new measure may shed new light on an old phenomenon. However, you should evaluate its reliability and validity. This may mean testing reliability and validity before you use your new measure in your research. Alternatively, you can use your measure in your research and demonstrate its reliability and validity based on your results. A danger with this latter strategy is that if the measure has problems with reliability or validity, the results of your research will be questionable. Because demonstrating the validity, reliability, and accuracy of a new measure can be time consuming and expensive, using measures that are already available (especially if you are new to a research area) is advisable.

## QUESTIONS TO PONDER

---

1. What is the reliability of a measure?
2. How does the concept of reliability apply to different types of measures?
3. What is meant by the accuracy of a measure?
4. How do the reliability and accuracy of a measure affect the generality of the results of a study?
5. What is the validity of a measure?
6. What are the ways you can assess the validity of a measure?
7. What is the relationship between the reliability and validity of a measure?

## Scale of Measurement

The phrase *scale of measurement* usually refers to the units in which a variable is measured: centimeters, seconds, IQ points, and so on. However, this phrase also can refer to the *type* of scale represented by a given set of units. Stevens (1946) identified four basic scales, which can be arranged in order of information provided about the values along the scale. These are the nominal, ordinal, interval, and ratio scales. Stevens argued that the type of scale along which a given variable is measured determines the kinds of statistical analyses that can be applied to the data. Because some kinds of statistical analysis are more informative and sensitive than others, it is important that you carefully consider the scale of measurement when evaluating the suitability of a given variable for your study. You should learn the characteristics of each scale and be able to identify the type of scale a given variable represents.

**Nominal Scales** At the lowest level of measurement, a variable may simply define a set of cases or types that are qualitatively different. For example, sex may be male or female. According to one scheme, a person's personality may be classified as introverted or extraverted. Variables whose values differ in quality and not quantity are said to fall along a **nominal scale**. In a nominal scale, the values have different names (in fact, the term *nominal* refers to name), but no ordering of the values is implied. For example, to say that male is higher or lower in value than female makes no sense. They are simply different.

Sometimes the qualitative values of a nominal-scale variable are identified by numbers (typically for the purpose of computer analysis). For example, three candidates for political office—Smith, Jones, and Brown—might be assigned the numbers 0, 1, and 2, respectively. If the assignment of numbers to the different qualitative values is arbitrary and does not imply any quantitative ordering of the values, then the results of certain mathematical calculations on these numbers will be meaningless.

To see that this is true, imagine that you determine how many voters voted for Smith, for Jones, and for Brown in a recent election and identify each candidate by a number as indicated above. You compute the average vote, which turns out to be 1.5. What does it mean to say that the average vote was 1.5? Does it mean that the average voter favored a candidate who was halfway between Jones (Candidate 1) and Brown (Candidate 2)? That seems doubtful. Moreover, had you assigned different numbers to the three candidates (say, Brown = 0, Smith = 1, and Jones = 3), you would have obtained a different average.

Although it makes no sense to apply mathematical operations to nominal values (even when these values are represented by numbers), you *can* count the number of cases (observations) falling into each nominal category and apply mathematical operations to those counts. So, you could count the number of voters who cast their ballots for Smith, Brown, and Jones and see which candidate garnered the most or least number of votes. Such counts fall along a ratio scale (see below).

**Ordinal Scales** At the next level of measurement are variables measured along an **ordinal scale**. The different values of a variable in an ordinal scale not only have different names (as in the nominal scale) but also can be ranked according to quantity. For example, a participant's self-esteem may be scored along an ordinal scale as low, moderate, or high. However, the distance between low and moderate and between moderate and high is not known. All you can say for sure is that moderate is greater than low and high is greater than moderate. Because you do not know the actual amount of difference between ordinal values, mathematical operations such as addition, subtraction, multiplication, and division, which assume that the quantitative distance between values is known, are likely to produce misleading results. For example, if three teams are ranked first, second, and third, the difference in ranking between first and second and between second and third are both 1.0. This implies that the teams are equally spaced in terms of performance. However, it may be the case that the first- and second-place teams are almost neck-and-neck and both performing far above the third-place team.



**Interval and Ratio Scales** If the spacing between values along the scale is known, then the scale is either an **interval scale** or a **ratio scale**. In either case, you know that one unit is larger or smaller than another, as well as by how much.

The two types of scale differ as follows. A ratio scale has a zero point that literally indicates the absence of the quantity being measured. An interval scale has a zero point that does *not* indicate the absence of the quantity. With interval scales, the position of the zero point is established on the basis of convenience, but its position is purely arbitrary.

The Celsius scale for temperature is an interval scale. Its zero point does not really indicate the absence of all temperature. Zero on the Celsius scale is the temperature at which ice melts—a convenient, easy-to-determine value. Although this temperature may seem cold to you, things can get much colder. In contrast, the Kelvin scale for temperature is a ratio scale. Its zero point is the temperature at which all heat is absent. You simply can't get any colder.

In psychological research, when you measure the number of responses on a lever in an operant chamber, you are using a ratio scale. Zero responses means literally that there are no responses. Other examples of psychological research data measured on a ratio scale are the number of items recalled in a memory experiment, the number of errors made in a signal-detection experiment, and the time required to respond in a reaction-time experiment. Again, zero on these scales indicates an absence of the quantity measured. In contrast, if you have participants rate how much they like something on a scale from 0 to 10, you are using an interval scale. In this case, a rating of zero does not necessarily mean the total absence of liking.

For practical purposes, an important difference between interval and ratio scales concerns the kinds of mathematical operations that you can legitimately apply to the data. Both scales allow you to determine by how much the various data points *differ*. For example, if one participant makes 30 responses and a second makes 15 responses, you can confidently state that there is a 15-response difference between participants. If the data are measured on a ratio scale (as in this example), you can also correctly state that one participant made half as many responses as the other (i.e., you can divide one quantity by the other to form a ratio). Making ratio comparisons makes little sense when data are scaled on an interval scale. Consider the IQ scale of intelligence, which is an interval scale. If one person has an IQ of 70 and another an IQ of 140, saying that the person with the 140 IQ is twice as intelligent as the person with the 70 IQ is nonsense. The reason is that even a person scoring zero on the test may have some degree of intelligence.

### Variables and Scales of Measurement

The four basic scales of measurement identified by Stevens (1946) help clarify the level of information conveyed by the numbers that result from measuring some variable. However, they should be viewed only as rough guides to aid in thinking about the numbers. Not all measures fall precisely into one or the other scale category; for example, many psychological measures do not seem to fall along a scale of precisely equal intervals, as required of an interval-scale measure, yet the distances between values along the scale are known with greater precision than

would be implied by the mere rank ordering of an ordinal scale. Researchers usually analyze such measures as if they had full interval-scale properties. Furthermore, it is possible to construct alternatives or additions to the basic scales. For example, Mosteller and Tukey (1977) offer an alternative classification that includes seven categories: amounts, counts, counted fractions (ratios with a fixed base, such as “8 out of 10 doctors”), names (categories with no particular order), ranks, grades (categories with a natural order), and balances. This scheme is based on the nature of the values rather than on what logical or mathematical operations legitimately can be performed on them.

Despite these caveats, Stevens’s (1946) four basic scales do at least highlight the information content of a set of numbers representing some particular variable as measured. In the next section, we discuss several factors that you should consider when deciding on a scale of measurement to adopt for some variable to be included in your study.

### Choosing a Scale of Measurement

You should consider at least three factors when choosing a scale of measurement for a given variable: the information yielded, the statistical measures that you would like to apply to the data, and, if you expect to apply your results directly to natural settings, the ecological validity of the measure.

**Information Yielded** One way to think about the four scales of measurement described is in terms of the amount of information that each provides. The nominal scale provides the least amount of information: All you know is that the values differ in *quality*. The ordinal scale adds crude information about *quantity* (you can rank the order of the values). The interval scale refines the measurement of quantity by indicating how much the values *differ*. Finally, the ratio scale indicates precisely how much of the quantity exists. When possible, you should adopt the scale that provides the most information.

**Statistical Tests** As noted, Stevens (1946) argues that the basic scale of measurement of a variable determines the kinds of statistics that can be applied to the analysis of your data. Typically, the statistics that are used for nominal and ordinal data are less powerful (i.e., less sensitive to relationships among variables) than are the statistics used for interval or ratio data. (See Chapter 14 for a more detailed discussion of the power of a statistical test.) On a practical level, this means that you are less likely to detect a significant relationship among variables when using a nominal or an ordinal scale of measurement.

Many statisticians now believe that this view is overly restrictive. They suggest that the numbers resulting from measurement are just numbers and that a statistical analysis does not “care” how the numbers were derived or where they came from (e.g., see Velleman & Wilkinson, 1993). To illustrate this viewpoint, Lord (1953) tells a story about football jerseys being sold to the football team on campus. Each jersey displayed a number. When used to identify which jersey belongs to whom, the numbers serve only as names; they might just as well be letters when used for this

purpose, and thus they represent a nominal scale of measurement. However, according to the story, after quite a number of jerseys had been sold, the members of the freshman team became quite unhappy when the sophomore team began laughing at them because the freshman players' jerseys all had low numbers. The freshman players suspected that a trick was being played on them, so they asked a statistician to investigate. The statistician immediately computed the mean (average) jersey number for the freshman players for all the jerseys that had been in the store's original inventory. The freshman students were indeed getting more than their fair share of low numbers, and the probability that this was a chance event was so low as to be, for all practical intents, zero.

To compute the means, the statistician used the jersey numbers as quantities along an interval scale of measurement as if larger numbers indicated larger "amounts" of some variable. Indeed, both freshmen and sophomores were behaving as if the numbers represented something like social status, with low numbers corresponding to low status and high numbers corresponding to high status. In fact, the analysis in terms of means was meant to discover whether the jersey numbers were systematically assigned according to class rank rather than being arbitrarily assigned substitutes for the player's names, as would normally be the case for nominally scaled values.

As Lord's story makes clear, the scale of measurement that applies to a number depends on how the number is to be interpreted. However, in most cases, you know when designing the study how you would like to go about analyzing the data and therefore what assumptions the data will have to meet when you apply those analyses to them. For example, computing means of numbers representing a set of three or more nominal-scale categories would make no sense because the values of those means would change depending on which numbers were used to identify which categories.

**Ecological Validity** The discussion thus far would indicate that you should use ratio or interval scales whenever possible in order to maximize the amount of information contained in the data. However, your research question may limit your choice of a measurement scale. If you are planning to conduct applied research, for example, you may be forced to use a certain scale even if that scale is one of the less informative ones. Consider the following example.

One author of this book (Bordens, 1984) conducted a study of the factors that influence the decision to accept a plea bargain. In this study, participants were told to play the role of either an innocent or a guilty defendant. They then were given information concerning the likelihood of conviction at trial and the sentences that would be received on conviction at trial or after a plea bargain.

In this situation, the most realistic dependent measure is a simple "acceptance–rejection" of the plea bargain. Real defendants in plea bargaining must make such a choice, so a dichotomous accept–reject measure was used even though it employs a less informative (dichotomous) scale of measurement (nominal). Sometimes you must compromise your desire for a sensitive measurement scale so that you will have an ecologically valid dependent measure (Neisser, 1976). A dependent measure has *ecological validity* if it reflects what people must do in real-life situations.

Adopting a more limited (nominal, ordinal, dichotomous) scale for your measure (even if it results in an ecologically valid measure) has two problems: The amount of information is limited, and the statistical tests that can be applied are less powerful. If you need to adopt a more limited measure to preserve ecological validity, you may be able to circumvent the limitations of scale by using special techniques.

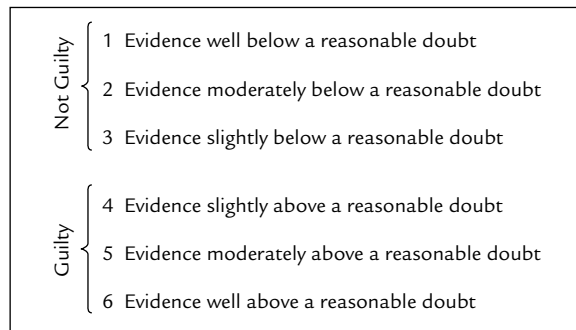
One technique is to include an interval or ratio scale in your study along with your nominal or ordinal measure. Before you analyze your data, you can create a *composite scale* from these measures. A composite scale is one that combines the features of more than one scale.

In the plea-bargaining study, Bordens (1984) included both a nominal dichotomous accept–reject measure and an interval scale (participants rated how firm their decisions were on a scale ranging from 0 to 10). A composite scale was created from these two scales by adding 11 points to the firmness score of participants who rejected the plea bargain and subtracting from 10 the firmness scores of participants who accepted the plea bargain. The resulting scale (0 to 21) provided a continuous measure of degree of firmness of a participant’s decision to accept or reject a plea bargain (0 was *firmly accept*, and 21 was *firmly reject* the plea bargain). The composite scale was reported along with the dichotomous accept–reject measure. The composite scale revealed some subtle effects of the independent variables that were not apparent with the dichotomous measure.

Another strategy you can use when you feel that a dichotomous scale is important is to arrange an interval scale so that a dichotomous decision is also required. For example, Horowitz, Bordens, and Feldman (1980) developed a scale that preserved some of the qualities of an interval scale while yielding dichotomous data. To assess the guilt or innocence of a defendant in a simulated criminal trial, Horowitz et al. used the 6-point, bracketed scale illustrated in Figure 5-1. Notice that points 1 through 3 are bracketed as a not-guilty verdict and points 4 through 6 are bracketed as a guilty verdict. The points on the scale were labeled so that participants could also rate the degree to which the evidence proved either guilt or innocence. This scale forced participants to decide that the defendant was either guilty or innocent while yielding a more sensitive measure of the effects of the independent variables.

**FIGURE 5-1** A bracketed 6-point scale.

SOURCE: Based on Horowitz, Bordens, and Feldman, 1980.



## QUESTIONS TO PONDER

---

1. What are the defining characteristics of Stevens's four scales of measurement? Do all measures fall neatly into one of the four categories?
2. What factors affect your choice of a scale of measurement?
3. What is ecological validity, and why should you be concerned about it?

### Adequacy of a Dependent Measure

You might find that a carefully planned dependent measure looks better on paper than it works in practice. Two potential problems involve the sensitivity of the dependent measure and range effects.

**Sensitivity of the Dependent Measure** Some measures of a dependent variable may be insensitive to the effect of a manipulation, whereas other measures under the same conditions definitely show an effect, as was clearly demonstrated to one author of this book (Abbott) in a study designed to investigate the role of the cerebral cortex in the expression of fear. Normal laboratory rats and rats whose cortexes had been surgically removed immediately after birth were exposed to three brief (0.5-second), relatively mild foot shocks in an operant chamber, where they were observed for several minutes. During the observation period, fear of the chamber cues was assessed by recording the number of 2-second intervals during which the rats "froze" (remained immobile). The normal rats froze during most of the observation period (as is typical), but no freezing was observed in the decorticate rats.

If only observations of freezing had been collected, the experimenter would have concluded from these data that the shocks had absolutely *no* effect on the post shock behaviors of the decorticate rats. However, unsystematic observations made during the course of the experiment revealed that, far from being unaffected by the shock, the behaviors of the decorticate rats changed radically. Even with almost no freezing, exploratory activity (which had been going on strongly prior to shock) all but ceased following the shocks and was replaced by a tentative stretch-and-quickly-withdraw behavior. Although frequently observed prior to shock, standing on the hind legs alone was all but absent following shock.

Unfortunately, these behaviors were not carefully defined and systematically recorded. The experimenter could refer only to impressions of behavioral change rather than to hard data. To determine the precise effect of the shocks on decorticate behaviors, the experiment must be run again, with the dichotomously scaled freezing measure replaced by a ratio-scaled continuous measure of behavioral activity, and the incidence of other behaviors (such as rearing) must be recorded.

In this case, the measure of freezing was insensitive to the subtle changes in behavior brought about by the independent variable. This was the case despite the fact that the measure had proven effective in other experiments. Unsystematic observations carried out during the course of the experiment can provide a useful check on the adequacy of your measure and may reveal defects as they did here. Although you may have to redesign and rerun your study, your understanding of the phenomenon under investigation will benefit.

**Range Effects** In addition to worrying about the sensitivity of your dependent variable, you need to be concerned with what are commonly called range effects. **Range effects** occur when the values of a variable have an upper or lower limit, which is encountered during the course of the observation. Range effects come in two types: floor effects and ceiling effects. As you might expect from the names, floor effects occur when the variable reaches its lowest possible value, whereas ceiling effects occur when the variable reaches its highest possible value.

The problems that range effects can cause are subtle and pernicious (harmful). They are subtle in that you don't always know that you have encountered them. Range effects are pernicious in that their consequences are hard to deal with after the fact and may require a redesign of the study. Consider the following example.

Assume that you have decided to study the effect of retention interval on memory for fruit and vegetable words (you happen to be fond of salads). You settle on a set of retention intervals that span 10 to 100 minutes in 10-minute increments, and you decide to measure retention by having participants attempt to pick out the correct word from a list of 10 items. The retention score for each participant is the percentage of correct choices in 10 trials.

You vary the retention interval across trials and get a retention score for each interval. To your surprise, you find absolutely no effect of retention interval. Averaged across participants, retention is about 95% at each interval!

Fortunately, you are aware of the potential for range effects in your data and stop to examine the scores more closely before concluding that retention interval has no effect on memory for fruit and vegetable words. Looking at the scores of each participant, you realize that 19 out of 20 participants have scored perfectly at every interval. Could the retention task be too easy?

It is possible that differences in retention might have been detected if the task were more demanding. Perhaps there *is* an effect of retention interval on memory. In this case, however, even at the longest interval, memory was still good enough to score 100% correct on the retention task. Because 100% was the upper limit of your measure, showing any better retention at shorter intervals was impossible. You have encountered a ceiling effect.

Range effects affect your data in two distinct ways: First, by limiting the values of your highest (or lowest) data points, the range effect decreases the differences between your treatment means. The apparent effects of your independent variables are lessened, perhaps to the extent that no statistically reliable differences will surface between them. Second, the variability of scores within the affected treatments is reduced. Because many commonly used inferential statistics estimate variability due to random causes from the variability of scores within the treatments, these statistics tend to give misleading results. In this case, they will usually underestimate the probability of the observed differences in treatments arising through chance alone. (See Chapter 14 for a discussion of inferential statistics and how they work.)

Because range effects distort your data both in central tendency and in variability, do your best to avoid them. Previous research often provides a guide, but on some occasions you may need to determine appropriate methods by trial and error.

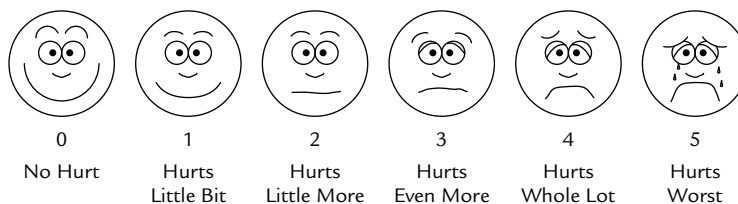
## Tailoring Your Measures to Your Research Participants

As another aspect to designing appropriate measures, you must consider the capabilities of your research participants. If you are working with young children or mentally impaired adults, you must tailor your measure to their level of understanding. It makes little sense to use a complicated rating scale with complex instructions if your participants have limited mental capacities.

One way to tailor the dependent measure to your participants is to represent your measures graphically. For example, instead of using a rating scale to measure a preference among young children (perhaps for a toy), you could use a more concrete measure. The child could be asked to give you a number of blocks, blow up a balloon, or vary the space between two objects to indicate the degree of preference. Another technique used with children is to adapt rating scales to a visual format. For example, a scale for children to rate pain that they are experiencing uses a series of six cartoon faces with varying expressions (Wong & Baker, 1988). Children point to the face that best reflects the amount of pain they are experiencing. This scale is shown in Figure 5-2. Cartoon faces could also be used to represent the points on a rating scale. Creative measurement techniques also may be needed when dealing with intellectually impaired or very old adults.

Some good examples of creative measurement techniques are those developed to study infant development. With preverbal infants, you have the problem that the participants of your study cannot understand verbal instructions or respond to measures as would an older child or an adult. Consequently, researchers of infant behavior have developed techniques to indirectly test the capabilities of the infant. Popular techniques used with preverbal infants include habituation, preference testing, and discrimination learning. The *habituation technique* capitalizes on the fact that even infants get bored with repeatedly presented stimuli. For example, in a study of the ability to discriminate shapes, you might repeatedly present the infant with a square until the infant no longer looks at the stimulus. You would then present a new stimulus (a circle). If the infant looked at the circle, you could infer that the infant could tell the difference between the two stimuli.

Alternatively, you could investigate the same problem with the *preference technique*. Here you present the two stimuli simultaneously. If the infant looks at one stimulus more than the other, you can infer that the infant can distinguish them.



**FIGURE 5-2** The Wong-Baker faces pain rating scale

SOURCE: <http://intelihealth.com/IH/ihtIH/WSIHW000/29721/32087.html#wong>; reprinted with permission.

In *discrimination learning*, you attempt to train different behaviors to the different stimuli (e.g., suck when a square is present, but not a circle). Differential rates of responding suggest the capacity to discriminate.

The need to tailor a measure to your participants is not limited to children and impaired adults. Even adults of normal intelligence may have difficulty responding to your measures. Remember that your participants are probably naive to the research jargon with which you are familiar. For example, they may not understand what you mean when you say that increasing numbers on a scale represent an increase in whatever is being studied. Whenever you suspect that your participants may misunderstand how to use the measure, make a special effort to clearly describe it. For example, Figure 5-3 shows how a scale from 0 to 10 can be graphically presented. Notice how the arrow increases in width as the numbers increase. Such a visual presentation may help participants understand that a 7 means they feel more strongly and a 4 less so.

Regardless of the measure chosen, pretest it to ensure that it is appropriate for your participants. During the pretest, you may find that your measure must be modified to fit the needs of your research. Such modifications can then be made before you invest large amounts of time and effort in your actual study.

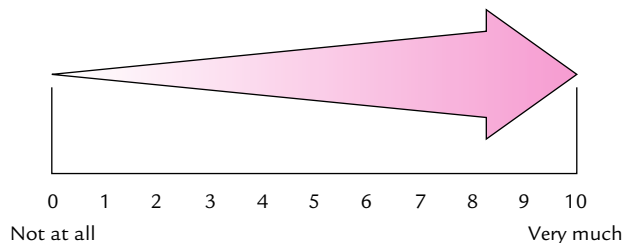
## QUESTIONS TO PONDER

---

1. What is meant by the adequacy of a dependent measure?
2. What is meant by the sensitivity of a dependent measure, and why should you be concerned about it?
3. What are range effects, and why should you be concerned about them?
4. When should you consider tailoring your dependent measures to the needs of your research participants?
5. How can you tailor your dependent measures? (Give examples.)

## Types of Dependent Variables and How to Use Them

Now that we have covered some of the basics of measurement and scaling, we can examine the types of dependent variables in psychological research and their uses.



**FIGURE 5-3** How to format a rating scale to reinforce the idea that increasing numbers represent an increasing amount of some characteristic.



The following sections describe four types of dependent measures: behavioral measures, physiological measures, self-report measures, and implicit measures.

**Behavioral Measures** Although the number of dependent variables is potentially vast, those used in behavioral research do tend to fall into a few basic categories. One type of dependent measure is a **behavioral measure**. When using a behavioral measure, you record the actual behavior of your subjects. In a study of helping behavior, for example, you might expose participants to different treatments (such as having a male or a female experimenter drop some packages) and then take note of the behavior of your participants (such as whether or not a participant helps).

One behavioral measure is the *frequency* of responding. To determine the frequency of a behavior, you count the number of occurrences over some specified period. For example, Goldiamond (1965) calculated the frequency of stuttering in a behavior modification study. Participants read pages of text, and Goldiamond counted the instances of stuttering across successive pages. Goldiamond found that the rate of stuttering declined during periods in which stuttering was punished with bursts of loud noise. Frequency counts also can be made over successive time periods.

Another behavioral measure is *latency*. Here you measure the amount of time it takes for subjects to respond to some stimulus. In the helping experiment described earlier, you could have measured how long it took participants to offer help in addition to whether or not participants helped. Any measure of reaction time is a latency measure.

In some types of research, *number of errors* might be an appropriate behavioral measure, which can be used with a well-defined “correct” response. Learning experiments often record number of errors as a function of the number of learning trials.

Behavioral measures are fine indicators of overt behavior. However, with a behavioral measure, you may not be able to collect data dealing with the underlying causes for behavior. To gain insight into the factors that underlie behavior, you often must follow up behavioral measures with other measures.

**Physiological Measures** A second type of dependent variable is a **physiological measure**. This type of measure typically requires special equipment designed to monitor the participant’s bodily functions. Such measures include heart rate, respiration rate, electrical activity of the brain, galvanic skin resistance, and blood pressure, among others.

A good example of the application of physiological measures is found in research on sleep. Participants come to the sleep laboratory, and physiological responses such as brain activity (measured with an electroencephalogram, or EEG), heart rate, respiration rate, and eye movements are recorded. This research has shown that the activities of the brain and body change cyclically over the course of a night’s sleep.

Modern brain-imaging techniques such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) have opened a new window into the dynamic activity of the brain during various kinds of mental tasks and have highlighted differences in brain functioning between normal individuals and those diagnosed with mental disorders such as schizophrenia.

The physiological measures just described are all noninvasive and do not harm the participant. Invasive measures, which sometimes do inflict a degree of harm, usually require the use of animals. For example, a physiological psychologist may implant an electrode into a rat's brain in order to record the activity of particular brain cells while the animal learns to perform a new behavior. Changes in brain-cell activity during learning constitute the dependent variable.

Physiological measures provide you with fairly accurate information about such things as the state of arousal within the participant's body. A drawback to this type of measure is that you often must infer psychological states from physiological states. As noted in Chapter 1, whenever you make inferences, you run the risk of drawing incorrect conclusions.

**Self-Report Measures** A third method commonly used to assess behavior is the **self-report measure**. Self-report measures take a variety of forms. One common form is the *rating scale*. In a study of jury decision making, for example, participants could rate the degree of guilt on a scale ranging from 0 to 10. A popular method in attitude assessment is *Likert scaling*. Participants are provided with statements (e.g., "Nuclear power plants are dangerous") and are asked to indicate the degree to which they agree or disagree with the statement on a 5-point scale ranging from 1 (strongly disagree) to 5 (strongly agree). (See Chapter 9 for more information on rating scales.)

Rating scales are but one method of quantifying a dependent variable. Another method is **Q-sort methodology**, a qualitative measurement technique that involves establishing evaluative categories and sorting items into those categories. The method, pioneered by William Stephenson in 1935, is a technique for exploring subjectivity in a wide variety of situations (Brown, 1996). For example, if you are interested in having participants evaluate poems representing different literary styles, you can have participants read short poems printed on index cards and then sort them into seven evaluative categories: dislike very much, dislike somewhat, dislike slightly, neutral, like slightly, like somewhat, and like very much. This process is repeated with a number of participants. You can then analyze the data from each participant to determine whether any significant patterns exist (e.g., a general liking for Haiku and a general disliking for blank verse). This can be accomplished with specialized Q-sort correlational and factor analytic techniques (see Brown, 1996, for an example). You can also enter Q-sort data into a standard analysis of variance to explore main effects and interactions among variables.

Self-report measures are highly versatile and relatively easy to use. You can ask participants to evaluate how they are feeling at the present time. In the jury decision example, participants would be providing an evaluation of the defendant's guilt immediately after exposure to a trial. In other cases, you may ask participants to reflect on past experiences and evaluate those experiences. This is referred to as a *retrospective verbal report* (Sheridan, 1979). In still other cases, you may ask for a *prospective verbal report* (Sheridan, 1979). Here you would ask participants to speculate on how they *would* react in a certain future situation.

Although self-reports are popular and relatively easy to use, they do suffer from reliability and validity problems. When using the retrospective verbal report, you must be aware of the possibility that the measure is somewhat invalid. You cannot really be sure that the participant is giving an accurate assessment of prior behaviors. The participant may be giving you a report or reconstruction of how he or she felt about the behavior that you are studying rather than a true account of what happened. The report provided by the participant could be clouded by events that intervened between the original event and the present report. Validity is lowered to the extent that this report is at variance with the actual behavior. Similarly, prospective verbal reports require a participant to speculate about future behavior. In this case, you cannot be sure that what the participant says he or she will do is what he or she actually does. For these reasons, a self-report measure should be used along with another measure whenever possible.

Another problem with self-report measures is that you cannot be sure that participants are telling the truth. Participants have a tendency to project themselves in a socially desirable manner. In a study of racial prejudice using a Likert-scaling technique, for example, participants may not be willing to admit that they have prejudicial attitudes. In fact, research in social psychology has found that self-reports of attitudes (especially on sensitive topics) often do not accurately reflect actual attitudes. You can detect responses that project social desirability by including questions that, if the participant agrees (or disagrees) with them, indicate self-effacement (such as “I have never had a bad thought about a member of a racial minority”). If a participant says he or she has *never* had such a thought, he or she is probably responding in a socially desirable way.

**Implicit Measures** A dependent measure that has become increasingly popular in social psychology to measure attitudes and prejudice is an implicit measure. An *implicit measure* measures responses that are not under direct conscious control. For example, a person who is prejudiced may not admit to being prejudiced on a self-report measure. However, the person may show an emotional reaction to a person of a given race. An experiment by Correll, Park, Judd, and Wittenbrink (2002), in which participants playing a video game had to decide instantaneously whether to shoot or not shoot a potentially armed suspect, used an implicit measure of prejudice: the difference in likelihood that a Black or White suspect would be shot.

The most popular measure of implicit attitudes is the **Implicit Association Test (IAT)** developed by Greenwald, McGhee, and Schwartz (1998). In the IAT, you are presented with a set of words or images that you classify into groups (e.g., good/bad; Muslim/other person) as quickly as you can. The theory behind the measure is that you will more quickly associate positive characteristics (e.g., smart and happy) with members of a social group that you like than with one that you dislike. Because stimuli are presented rapidly and you are instructed to respond as rapidly as possible, your responses are generally outside of your conscious control. Results from studies that use this measure often find that even those who say they are not prejudiced show a preference for one group over another on the IAT. (You can try out the IAT for yourself online at <https://implicit.harvard.edu/implicit/>.)

## QUESTIONS TO PONDER

---

1. What are defining characteristics of the four types of dependent variables?
2. What are the advantages and disadvantages of each?
3. What is Q-sort methodology, and when is it used?
4. What do implicit measures reveal?

## CHOOSING WHEN TO OBSERVE

---

After you have chosen what to observe and how to measure it, you need to decide when you will make your observations. If you are performing laboratory research, experimental sessions generally are *when* you would observe. However, even within experimental sessions, you must still decide when observations are to be made.

As with the other aspects of your design, *when* you observe may be determined by established practices. For example, if previous research has proven the adequacy of a time-sampling procedure (you make observations at 10-minute intervals during the session), then making continuous observations may be safely abandoned in favor of the less demanding technique.

Your decision of when to observe may have to take into account the resources that you have at your disposal, particularly if the required observations must be made frequently or over long periods of time. For example, in the “freezing” experiment described earlier, an enormous amount of time would be required to code freezing behavior across consecutive 2-second intervals of time during an experimental session that lasted 5 hours, especially if a large number of subjects were to be observed. In such cases, you may be able to adopt a sampling strategy and make occasional observations at randomly chosen intervals during the session. Averaged over a number of subjects, such observations could provide a representative picture of changes across time.

An even better solution than the sampling strategy is to automate the observations. For example, Robert Leaton and George Borszcz of Dartmouth College describe a way to automatically record the freezing behavior of rats (Leaton & Borszcz, 1985). They suspended the observation chamber between stiff springs. A bar magnet affixed to the chamber moved slightly up or down whenever the rat made the slightest move, generating an electric current in a coil of wire through which the magnet passed. When the rat froze, movement ceased and the current disappeared. A microcomputer counted the passing intervals of time and scored each interval for the presence or absence of movement. With the apparatus used by Leaton and Borszcz, it was possible to continuously observe the freezing behavior during sessions of any desired length.

Of course, it is important to show that any device provides a good measure of the variable before you adopt the measure. One good measure of success is the degree to which the new measurements agree with measurements done “the old-fashioned way.” In the case of the automated freezing measure, Leaton and Borszcz (1985) demonstrated that the automatic readings correlate highly with personal observations. Techniques for automating your experiment are discussed in more detail later in the chapter.

## THE REACTIVE NATURE OF PSYCHOLOGICAL MEASUREMENT

---

One advantage that physicists and chemists have over psychologists when it comes to conducting research is that the “subjects” of physical and chemical experiments (e.g., balls rolling down inclined planes) pay absolutely no attention to the fact that they are participants in an experiment. They behave as they ordinarily do in nature. The subjects (rats, pigeons) and participants (college students, human adults) of psychological research do pay attention to their status as such and may modify their behavior as a result of their perceptions. This “reactive” nature of subjects and participants must be considered when designing and assessing psychological research. This section describes the kinds of reactions, the situations that sometimes give rise to them, and the things you can do to minimize (or at least assess) their impact on your data. A discussion of research with human participants begins this section, followed by a discussion of research with animal subjects.

### Reactivity in Research with Human Participants

Assume for the moment that you have defined your population of participants and are now ready to acquire your participants and run your experiment. You plan to have volunteers sign up and come to your laboratory for your experiment. What can you expect from these creatures that we call human participants?

One thing to realize is that the psychological experiment is a social situation. You as the experimenter, by definition, are in a position of power over the participant. Your participant enters this situation with a social history that may affect how he or she responds to your manipulations. Assuming that the participant is a passive recipient of your experimental manipulations is a mistake. The participant is a living, thinking human being who will generate personal interpretations of your experiment and perhaps guide behavior based on these interpretations. In short, the participant is a reactive creature. The behavior that you observe in your participants may not be representative of normal behavior simply because you are making observations.

To help you understand the reactions of research participants to your experiment, imagine that you have volunteered for a psychological experiment for the first time. You are a first-year student enrolled in introductory psychology who has had a little experience with psychological research. As you sit waiting to be called for the experiment, you imagine what the experiment will be like. Perhaps you have just talked about Milgram’s obedience research in your psychology class or saw a documentary about it on television and are wondering if you are going to be given electric shocks or if the researcher is going to be honest with you. You wonder if you are going to be told the experiment is about one thing when it is actually about something else.

At last, the experimenter comes out of the laboratory and says she is ready for you. You are led into a room with a white tile floor, white walls, a stainless steel sink in the corner, some ominous-looking equipment in another corner, and rather harsh fluorescent lighting. The experimenter apologizes for the cold surroundings and says that she is a graduate student and had to settle for one of the animal labs to run her master’s thesis research. (Do you believe her?) You take a look around the room and

muster enough courage to ask whether you are going to be shocked. The experimenter chuckles and assures you that the experiment deals with memory for abstract pictures. She then begins to read you the instructions. At that moment, some workers begin to hammer out in the hall. (Is this part of the experiment?) Again, the experimenter apologizes. She explains that they are installing a new air-conditioning system in that wing of the building. You think you detect a hint of a smile on her face. You don't believe her. You have decided that the experimenter is really trying to test how well you can perform a memory task under distracting conditions. You decide to "show the experimenter" that you can do well despite her obvious attempt to trick you. The experimenter runs you through the experiment. Of course, you try your hardest to get all the items right.

After you have completed the memory test, the experimenter asks you whether you have any questions. You smugly tell her that you saw through her obvious deception and worked even harder to get the items correct. After all, you weren't born yesterday! To this the experimenter incredulously assures you that the noise was not part of the experiment and tells you that she will have to throw out your data. The experiment has been set back an entire day.

### Demand Characteristics

Consider the psychological experiment in the light of this example. As stated, the human participant in a psychological experiment does not passively respond to what you may expose him or her to. On entering the laboratory, your participant probably assesses you and the laboratory (Adair, 1973). Given these assessments, the participant begins to draw inferences concerning what the experiment is about.

The cues provided by the researcher and the context that communicate to the participant the purpose of the study (or the expected responses of the participant) are referred to as **demand characteristics**. Participants gain information about the experiment from these demand characteristics. Unfortunately for you, the participant may be paying attention to cues that are irrelevant to the experiment at hand (as happened in the previous example when you believed the noise created by the work crew was related to the experiment). With the information obtained from the demand characteristics, the participant begins to formulate hypotheses about the nature of the experiment (such as "The experiment is measuring my ability to perform under adverse conditions") and begins to behave in a manner consistent with those hypotheses. Problems occur when the participant's hypotheses differ from the intended purpose of the experiment. Adair (1973) refers to this class of demand characteristic as "performance cues."

A second source of demand characteristics centers on the participant. According to Adair (1973), a class of demand characteristics known as **role attitude cues** may signal the participant that a change in the participant's attitude is needed to conform to his or her new role as a research participant (Adair, 1973, p. 24). Further, Adair points out that participants enter experiments with preexisting attitudes that dispose them to react in either a positive or a negative way to the experimental manipulations. Through various demand characteristics, the experiment can cause the participant to change his or her attitudes (Adair, 1973, p. 26). Adair lists the following

three categories of predisposing attitudes of participants: the cooperative attitude, the defensive or apprehensive attitude, and the negative attitude.

The *cooperative attitude* is characterized by a strong desire to please the experimenter. According to Adair, volunteering for an experiment “seals a contract between the experimenter and the participant, fostering cooperative behavior” (1973, p. 26). Reasons for the cooperative attitude include a desire to help science, a desire to please the experimenter, a desire to perform as well as possible, and a desire to be positively evaluated by others.

Several demonstrations of the impact of this positive attitude on the outcome of an experiment have been made. For example, Orne (1962) demonstrated that participants will engage in a boring, repetitive task for hours to please the experimenter. Participants were provided 2,000 sheets of paper (on which were columns of numbers to add) and a stack of index cards (on which instructions were printed). They were instructed to select the first card (which told the participants to add the numbers on the page) and then to select the next card. The next card told the participants to tear the completed sheet into pieces (not less than 32) and then select another sheet and add the numbers. The cycle of adding numbers and tearing sheets continued for as long as a participant was willing to go on.

If you were a participant in this experiment, what would you do? You may have said, “I’d do it for a few times and then quit.” In fact, quite the opposite happened: Participants continued to do the task for hours. Evidently, participants perceived the test as one of endurance. The participants’ cooperative attitude in this example interacted with the demand characteristics to produce some rather bizarre behavior. This “good participant” effect was also shown in an experiment by Goldstein, Rosnow, Goodstadt, and Suls (1972).

Some participants enter the laboratory worried about what will happen to them and have an *apprehensive attitude*. One of us (Bordens) was conducting an experiment on jury decision making, and several participants, on entering the lab, asked if they were going to be shocked. This apprehension may stem from the participants’ perception of the experimenter as someone who will be evaluating the participants’ behavior (Adair, 1973). The apprehensive attitude also has been shown to affect the research outcome, especially in the areas of compliance and attitude change (Adair, 1973).

Some participants come to the laboratory with a *negative attitude*. Even though most participants are either positive or defensive (Adair, 1973), some participants come to the lab to try to ruin the experiment. This attitude was most prevalent when participants were *required* to serve in experiments. Required participation made many participants angry. The present rules against forced participation may reduce the frequency of negative attitudes. However, you cannot rule out the possibility that some participants will be highly negative toward the experiment and experimenter.

### Other Influences

In addition to demand characteristics and participants’ attitudes, evidence also indicates that events outside the laboratory can affect research. For example, Greene and Loftus (1984) conducted an experiment on jury decision making in which

eyewitness testimony was being studied. Around the time that the experiment was conducted and in the same city, a celebrated case of mistaken identification was being unmasked. Knowledge of that case was reflected in the data. Participants generally were more skeptical of the eyewitness in the study after finding out about the case than they were before. However, after a while, the impact of the case diminished, and the responses returned to “normal.”

The moral to this story is that participants are not passive responders to the experiment. The experiment is a social situation in which the interaction between participant attitudes and the experimental context may affect the outcome of the experiment. As a researcher, you must be aware of demand characteristics and take steps to avoid them or at least to assess their impact. As with other participant-related problems, demand characteristics, participant attitudes, previous research experience, and exposure to everyday life can affect both internal and external validity.

### The Role of the Experimenter

The participant is not the only potential source of bias in the psychological experiment. The experimenter can sometimes unintentionally affect the outcome of the experiment. Assume that you are running your first experiment, an experiment of your own design. Because you are a student, you will be testing your own participants.

You are sitting alone in your laboratory, awaiting the arrival of your first participant. You have butterflies in your stomach and are a bit apprehensive about how you will perform in the experiment. The experiment is important to you because it is required for a class that you need for graduation. At last, your first participant arrives, and you usher him into your laboratory. You begin to read your instructions (which you feel are well written) to your participant and are puzzled to see that your participant is obviously not understanding the instructions. However, you press on.

Your experiment deals with the ability of people to recall certain words embedded within the context of other words. You want to show that interference will occur when the words are embedded in a context of other similar words. You are going to read a list of words to your participant and then give a recall test. In the high-similarity condition, you unconsciously read the words at a faster rate than in the low-similarity condition. You notice later that your collected data consistently confirm your preexperimental hypothesis.

Now, analyze what has happened. You wrote your instructions, believing that your participants would be able to understand them. As it turns out, the instructions were less clear than you thought. The problem here was that you assumed too much about the ability of your participants to understand the instructions. This may happen because you are used to talking to other psychology majors or professors familiar with the jargon of your discipline. The participants may not have that advantage. One thing that you could do to detect this problem is to pretest the instructions.

**Experimenter Bias** In the classic 1960s sitcom *Mr. Ed*, Wilbur Post owned a horse named Ed with a special talent: Ed could talk. In each episode, Ed’s antics created some interesting problems for Wilbur. Over a half century before *Mr. Ed* hit the airwaves, another horse, named Hans, created a sensation in the entertainment world



in Europe. Hans, it seemed, could solve simple mathematical problems. His owner, Wilhelm Von Osten, took great pains to teach Hans to solve the problems and then took Hans on the road to entertain people. Von Osten would show Hans a card with a math problem (e.g., an addition problem), and Hans would begin clopping out the answer with his hoof. Hans would stop clopping his hoof when the correct answer was reached. Audiences were astounded and for two years Hans earned Von Osten a nice living.

Not everyone was taken with Hans's mathematical prowess. A scientist named Oskar Pfungst doubted that Hans was able to solve math problems. Instead, he believed that Hans was picking up subtle cues from Von Osten. So Pfungst designed a series of tests to see if Hans had the miraculous abilities claimed. In one test, Von Osten showed Hans a card with a problem. The catch was that Von Osten did not know what the problem was. In this and similar tests in which Von Osten was not allowed to see the problem being put to the horse, Pfungst found that Hans could not solve the problems. Pfungst believed that Hans was reading his trainer's behavior, looking for cues to signal when Hans should stop clopping his hoof. In fact, Pfungst found that as Hans began clopping, Von Osten would unconsciously tense up, which showed in his body position and facial expressions. As Hans reached the correct answer, Von Osten would unconsciously relax, signaling Hans that the correct answer had been reached (Wozniak, 1999). In cases where Von Osten did not know what problem Hans was to solve, Von Osten could not provide Hans with the unconscious signals, and Hans's performance deteriorated.

At this point, you may be asking yourself, "What does this have to do with my research? I don't plan on dragging a horse around to entertain people." Whoa! Let's slow down and see how the case of Hans relates to your research. In fact, it relates in a quite simple way. The "Clever Hans phenomenon" poses a potential threat to the validity of your research. Let's take a look at a modern-day research example to see how this might work.

There is a phenomenon known as *facilitated communication*, which involves a "facilitator" physically helping an impaired person communicate by touching letters on a screen. The facilitator supports the impaired person's hand while the person guides his finger to a symbol on a screen (Montee, Miltenberger, & Wittrock, 1995). Supposedly, this technique allows the impaired person to communicate with others in ways and at levels previously believed to be impossible. But is facilitated communication a real phenomenon or another example of the Clever Hans phenomenon? Let's find out.

Barbara Montee, Raymond Miltenberger, and David Wittrock (1995) conducted an experiment to test the validity of facilitated communication. Seven client-facilitator dyads participated in this experiment. The experiment was conducted in the client-facilitator pairs' normal setting (e.g., day-care center) at the usual time of day. The pairs completed several facilitated communication tasks involving describing an activity or naming a picture. The independent variable was the information provided to the facilitators prior to the facilitated communication session. In one condition (the known condition), the facilitator was informed about the activity that the client had engaged in or the picture that the client had been shown. In another condition (the unknown condition), the facilitator was not informed about the activity

or picture. In the final condition (the false feedback condition), the facilitator was given incorrect information about the activity or picture. The dependent variable was whether, using facilitated communication, the client correctly described the activity or named the picture.

The results from this experiment were rather dramatic and are shown in Table 5-1. As you can see, the client's ability to describe the activity or name the picture was almost totally dependent upon whether the facilitator had accurate information about the nature of the activity or picture. Just as Hans could not solve his math problems when Von Osten did not know the answer, so the clients in this experiment could not respond correctly unless the facilitators knew the answers.

In both the Clever Hans and facilitated communication situations, there was a common problem known as **experimenter bias** (Rosenthal, 1976). Experimenter bias creeps in when the behavior of the experimenter influences the results of the experiment. This influence serves to confound the effect of your independent variable, making it impossible to determine which of the two was responsible for any observed differences in performance on the dependent measure. Experimenter bias flows from at least two sources: expectancy effects and treating various experimental groups differently to produce results consistent with the preexperimental hypotheses.

When an experimenter develops preconceived ideas about the capacities of the participants, **expectancy effects** emerge. For example, if you believe that your participants are incapable of learning, you may treat them in such a way as to have that expectation fulfilled. Rosenthal (1976) reports a perception experiment in which the independent variable was the information provided to students acting as experimenters. Some students were told that, according to previous ratings, their participants should perform well. Others were told that the participants would probably perform poorly. The student experimenters also were told they would be paid twice as much if the results confirmed the prior expectations. Rosenthal reports that establishing the expectancy led to different behavior on the part of the participants in the two experimental groups. Rosenthal points out that such expectancy effects may be a problem in not only experimental research but also survey research and clinical studies. In the previous hypothetical example, you (as the experimenter) read the list of words to participants differently, depending on the condition to which they were assigned. If

**TABLE 5-1** Mean Number of Correct Responses Made in the Montee, Miltenberger, and Wittrock (1995) Experiment on Facilitated Communication

TASK	INFORMATION CONDITION		
	<i>Known</i>	<i>Unknown</i>	<i>False</i>
Picture naming	75	0	1.8
Activity identification	87	0	0

the experimenter knows what the hypotheses of the experiment are, he or she may possibly behave in a manner that leads participants into certain behaviors to confirm the hypotheses. Keep in mind that this could be quite unintentional. When running your own research, you may have a vested interest in the outcome of the study, particularly if you have developed a hypothesis that predicts a certain result. Consequently, your expectations may subtly influence the participants in the different groups.

These two sources of experimenter bias threaten both internal and external validity. If your behavior becomes a source of systematic bias or error, then you cannot be sure that the independent variable caused the observed changes in the dependent variable. External validity is threatened because the data obtained may be idiosyncratic to your particular influence.

Because experimenter bias can pose such a serious threat to internal and external validity, you must take steps to reduce the bias. You can do this by using a *blind technique* in which the experimenter and/or the subject is blind to (not aware of) what behavior is expected or what, if any, treatment the subject has been exposed to. In a **single-blind technique**, the experimenter does not know which experimental condition a subject has been assigned to. For example, in an experiment on the effect of children watching violent television on aggression, children could be randomly assigned to watch either violent cartoons or nonviolent cartoons. The measure of aggression could be the number of aggressive acts a child engages in during free play on the playground. In a single-blind experiment, the observers watching the children do not know the condition to which the children were assigned.

In some research situations, a **double-blind technique** is appropriate. Neither the experimenter nor the participants know at the time of testing which treatments the participants are receiving. If you were interested in testing the effects of a particular drug on learning abilities, for example, you would give some participants the active drug and some a placebo (perhaps an injection of saline solution). The participants would not know which treatment was being administered, thus reducing the possibility that the participants' expectations about the drug would affect the results. Furthermore, you would have an assistant mix the drugs and label them arbitrarily with some code, such as "A" and "B." As the experimenter, you would not be told which was the active drug and which was the placebo until after the experiment was completed and the data were analyzed. Thus, neither you nor the participant would know at the time of testing which treatment that the participant was receiving. Neither your expectations nor the participants' could systematically bias the results. This is the essence of the double-blind procedure.

Another method for reducing experimenter bias is to *automate* the experiment as much as possible. In a memory experiment, you could present your stimulus items and time their presentations accurately using a personal computer. The interval between stimulus presentations would be held constant, avoiding the possibility that you might present the stimuli more rapidly to one group than to the other. You also could automate the instructions by using a videotaped version of the instructions. All participants would thus be exposed to the same instructions. (Automation is more fully discussed later in this chapter.)

Other potential sources of experimenter bias include the sex, personality, and previous experience of the experimenter. It is beyond the scope of this book to explore

all the potential experimenter effects. See *Experimenter Effects in Behavioral Research* (Rosenthal, 1976) for a complete treatment of this topic.

### Reactivity in Research with Animal Subjects

The section on using human participants in research pointed out that the behavior of participants can be affected by the behavior of the experimenter and by demand characteristics. Similar effects can be found with animal subjects. For example, Rosenthal (1976) reports research in which experimenter expectancy influenced the rate at which animals learned to navigate a maze. Participants serving as experimenters were told that the rats they would be teaching to run a maze were either very bright (would learn the maze quickly with few errors) or very dull (would have trouble learning the maze). The animals were actually assigned at random to the experimenters. Rosenthal found that the animals in the “bright” condition learned more quickly than the animals in the “dull” condition. The differing expectations of the student experimenters led them to treat their rats differently, and these differences in treatment led to changes in the behaviors of the rats.

Use blind techniques to avoid these and other sources of experimenter bias in animal research. For example, in a study in which a drug is to be administered, the person making the observations of the animal’s behavior should not know which subjects received the drug and which received the placebo.

Remember that demand characteristics are cues that subjects use to guide behavior within an experiment. Although animals will not be sensitive to demand characteristics in the same way that human participants are, some features of your experiment may inadvertently affect your subject’s behavior. For example, you may be interested in how a rat’s learning capacity is affected by receiving electric shocks just before the opportunity to work for food. If you do not clean the experimental chamber thoroughly after each animal is tested, the animals may respond to the odor cues from the previous animal. These odor cues may affect the current animal’s behavior differently, depending on whether or not the previous animal had received a shock. You must remember that, much like the human participants, the animal subject is an active processor of information. Your animals sense cues that are not intended to be a part of the experiment and may behave accordingly. Ultimately, the internal validity of your experiment may be threatened by the effects of these cues.

### QUESTIONS TO PONDER

---

1. How can the act of measurement affect your subjects’ responses?
2. What are role attitude cues, and how might they affect the results of your study?
3. What are demand characteristics, and how can they affect the results of your study?

4. What is experimenter bias, and how can it affect the results of your study?
5. What measures can be taken to deal with reactivity in research?

## AUTOMATING YOUR EXPERIMENTS

---

Psychological research presents many opportunities for outside, uncontrolled variables to affect your results. Automation can help eliminate experimenter effects and increase the precision of your measures.

In addition, automation can save time. Automated equipment allows you to run subjects even if you cannot be present. This is most useful in animal research in which subjects can be left unattended in the testing apparatus. In this case, you simply start the testing program and then return at the end of the session to record the data and return the subjects to their home cages.

Automation has other advantages as well. Automated measurements tend to be more accurate and less variable because they are not subject to the vagaries of human judgment. An automated system is not likely to miss an important event because it was daydreaming at the moment or distracted by an aching back. Nor is such a system likely to misperceive what actually happened because of expectations about what will happen (eliminating this source of experimenter bias).

Conversely, automation can cause you to miss important details. The changes in behavior shown by the nonfreezing decorticate rats might not have been detected had the automated freezing measure of Leaton and Borszcz (1985) been in use. Even when all your measurements are automated, you should observe your subjects occasionally. What you learn from these observations may provide fruitful explanations for changes detected by your automated variables and may provide you with new ideas to test.

Techniques for automation include the use of videotaped instructions, timers to control the duration that a stimulus is present and to time the intervals between stimuli, and computers to control an experiment. Because the computer has become almost a standard piece of laboratory equipment, a brief discussion of its components and their uses is in order.

Relatively inexpensive personal computers can be programmed and outfitted with the hardware and software needed to fully automate your experiment. For example, a computer could be programmed to control complex schedules of reinforcement in an animal learning experiment. Computers can be used to control research conducted with humans as well as animals. For example, you could program your computer to present stimuli to be used in research areas such as human learning and memory, perception, developmental psychology, and decision making.

If you use computers to conduct your research, remember that the computer performs many of the more tedious tasks involved in your research quickly and accurately but always does what you tell it to do (even if you make a mistake). Your automated experiment will only be as good as your program. Whether you use commercially available software or programs that you write yourself, you must be intimately familiar with the commercial software or with the computer language that you will be using and know how to interface your computer with your equipment.

## DETECTING AND CORRECTING PROBLEMS

---

No matter how carefully you plan your study, problems almost inevitably crop up when you begin to execute it. Two methods you can use to minimize these problems and ensure the usefulness of the data you collect are conducting a pilot study and adding manipulation checks.

### Conducting a Pilot Study

A **pilot study** is a small-scale version of a study used to establish procedures, materials, and parameters to be used in the full study. Frequently, it is a study that began life as a serious piece of research but “went wrong” somewhere along the way. The decorticate rat study became a pilot study for this reason. However, many pilot studies are designed from the ground up as pilot studies, intended to provide useful information that can be used when the “real” study gets under way.

Pilot studies can save tremendous amounts of time and money if done properly. Perhaps you intend to conduct a large study involving several hundred participants in order to determine which of two methods of teaching works best in introductory psychology. As part of the study, you intend to hand out a large questionnaire to the students in several introductory psychology classes. Conducting a small pilot study (in which you hand out the questionnaire to students in only a couple of classes) may turn up inadequacies in your formulation of questions, inadequacies that lead to confusion or misinterpretation. Finding these problems *before* you train instructors in the two teaching methods, have them teach a full term, and then collect the questionnaires from 2,000 students is certainly preferable to finding the problems afterward.

Pilot studies can help you clarify instructions, determine appropriate levels of independent variables (to avoid range effects), determine the reliability and validity of your observational methods, and work the bugs out of your procedures. They also can give you practice in conducting your study so that you make fewer mistakes when you “do it for real.” For these reasons, pilot studies are often valuable.

You should also be aware of some negative aspects of pilot studies. Pilot studies require time to conduct (even if less than that of the formal study) and may entail some expenditure of supplies. Where animals are involved, their use for pilot work may be questioned by the local animal care and use committee (particularly if the procedures involve surgery, stressful stimulation, or deprivation). In these cases, you may want to use the best available information to determine procedures and to try to “get it right” the first time around. Then only if you guess wrong will the study become a pilot study.

## ADDING MANIPULATION CHECKS

---

In addition to the dependent measures of the behavior under study, you should include manipulation checks. A **manipulation check** simply tests whether or not your independent variables had the intended effects on your participants. They allow

you to determine if the participants in your study perceived your experiment in the manner in which you intended. For example, if you were investigating the impact of a person's attractiveness on how his or her work is evaluated, you might have participants evaluate an essay attributed to either an attractive or unattractive author. This could be done by attaching a photograph of an attractive or unattractive person to an author profile accompanying the essay. As a manipulation check, you could have participants rate the attractiveness of the author on a rating scale.

Manipulation checks also provide you with information that may be useful later when attempting to interpret your data. If your experiment yielded results you did not expect, it may be that participants interpreted your independent variable differently from the way you thought they would. Without manipulation checks, you may not be able to properly interpret surprising effects. Manipulation checks may permit you to determine why an independent variable failed to produce an effect. Perhaps you did not effectively manipulate your independent variable. Again, manipulation checks provide information on this.

A set of measures closely related to manipulation checks are those asking participants to report their perceptions of the entire experiment. Factors to be evaluated might include their perceptions of the experimenter, what they believed to be the true purpose of the experiment, the impact of any deception, and any other factors you think are important. Like manipulation checks, these measures help you interpret your results and establish the generality of your data. If you find that participants perceived your experiment as you intended, you are in a better position to argue that your results are valid and perhaps apply beyond the laboratory.

## QUESTIONS TO PONDER

---

1. What is a pilot study, and why should you conduct one?
2. What are manipulation checks, and why should you include them in your study?

## SUMMARY

---

In contrast to casual, everyday observations, scientific observations are systematic. Systematic observation involves making decisions about what, how, and when to make observations. Observations of behavior are made under controlled conditions using operational definitions of the variables of interest.

When choosing variables for your study, you should be guided in your choice by research tradition in the area of study, theory, the availability of new techniques or measures, and the limits imposed by the equipment available to you. In addition, you need to be concerned with the characteristics of the measure, including its reliability, its validity, and the level of measurement it represents. A measure is reliable to the extent that repeated measurements of the same quantity yield similar values. For measures of physical variables, reliability is indexed by the precision of the measure, and for population estimates, by the margin of error. The reliability of the judgments

of multiple observers is indexed by a statistical measure of interrater reliability. The reliability of psychological tests can be determined in a variety of ways, yielding test–retest, parallel-forms, or split-half reliabilities. A measure is accurate if the numbers it yields agree with a known standard. Accuracy is not a characteristic of most psychological measures because there are no agreed-upon standards for them. A measure is valid to the extent that it measures what it is intended to measure. Several types of validity assessment exist for psychological tests, including face validity, content validity, construct validity, and criterion-related validity. The latter takes two forms, called concurrent validity and predictive validity.

One aspect of systematic observation is developing dependent measures. Your data can be scaled along one of four scales of measurement: nominal, ordinal, interval, and ratio. Nominal and ordinal scales provide less information than do interval and ratio scales, so use an interval or ratio scale whenever possible. You cannot use an interval or ratio scale in all cases because some research questions demand that a nominal or ordinal scale be used. Your choice of measurement scale should be guided by the needs of your research question. When a less informational scale must be used to preserve ecological validity, you can preserve information by creating a composite scale from a nominal and interval scale. This will help you to “recover” information not yielded by a nominal scale.

Beyond choosing a scale of measurement, you must also decide how to design and collect your dependent measures. Your measures must be appropriate for your subject population. Consequently, you may have to be creative when you design your measures. You may count number of responses, which is a ratio scale. You can use interval scales in a variety of research applications. You must decide how to format these scales, how to present them to subjects, and how to develop clear and concise instructions for their uses.

In some research, your measure of behavior may be limited by range effects. That is, there may be an upper and lower limit imposed on your measure by the behavior of interest. For example, rats can run just so fast in a maze. Range effects become a problem when the behavior quickly reaches its upper or lower limit. In such cases, you may not detect a difference between two groups because of ceiling or floor effects. It is a good idea to conduct pilot studies to test your measures before investing the time and energy in your study. During the pilot study, you may find that your measures need to be modified.

There are four types of dependent variables you can use in your research: behavioral measures, physiological measures, self-report measures, and implicit measures. Behavioral measures include direct measures of behavior such as number of responses made or the number of errors made. Physiological measures involve measuring some biological change (e.g., heart rate, respiration rate, or brain activity). Physiological measures can be noninvasive (e.g., a PET scan) or invasive (e.g., implanting an electrode in a rat’s brain). Self-report measures have participants report on their own behavior and can be prospective (speculate on future behavior) or retrospective (report on past behavior). One special form of a self-report method is the Q-sort method in which participants classify stimuli into categories. Implicit measures measure unconscious reactions to stimuli and are used to tap into attitudes that individuals may not admit to overtly.



Observation in psychological research differs from observation in other sciences because the psychologist deals with living organisms. The participants in an experiment are reactive; they may respond to more in the experimental situation than the manipulated variables. Participants bring to the experiment unique histories and attitudes that may affect the outcome of your experiment.

Demand characteristics can be a problem in behavioral research. Participants pick up on cues from the experimenter and research context. These cues may affect the participant's behavior. Furthermore, the experimenter must be careful not to inadvertently affect the participants. Experimenter effects can be avoided by using blind techniques or automating your experiment or both. Automation can be done by videotaping instructions or applying computers to control your experiment or both.

Before conducting a study, it is a good idea to do a pilot study, which is a small-scale version of your study used to test the effectiveness of your materials, procedures, and other parameters. You can identify and correct problems before investing time and effort in your main study. It is also a good idea to include manipulation checks in your research. These are measures specifically designed to determine how your participants perceived the variables of your study. This information can help you to interpret your results and identify problems that may have emerged in your study.

## KEY TERMS

---

reliability

test-retest reliability

parallel-forms reliability

split-half reliability

accuracy

validity

face validity

content validity

criterion-related validity

concurrent validity

predictive validity

construct validity

nominal scale

ordinal scale

interval scale

ratio scale

range effects

behavioral measure

physiological measure

self-report measure

Q-sort methodology

Implicit Association Test (IAT)

demand characteristics

role attitude cues

experimenter bias

expectancy effects

single-blind technique

double-blind technique

pilot study

manipulation check

## 6

## CHAPTER

## CHAPTER OUTLINE

**General Considerations**

- Populations and Samples
- Sampling and Generalization
- Nonrandom Sampling
- Is Random Sampling Always Necessary?

**Acquiring Human Participants for Research**

- The Research Setting
- The Needs of Your Research
- Institutional Policies and Ethical Guidelines

**Voluntary Participation and Validity**

- Factors That Affect the Decision to Volunteer
- Volunteerism and Internal Validity
- Volunteerism and External Validity
- Remedies for Volunteerism

**Research Using Deception**

- Types of Research Deception
- Problems Involved in Using Deception
- Solutions to the Problem of Deception

**Considerations When Using Animals as Subjects in Research**

- Contributions of Research Using Animal Subjects
- Choosing Which Animal to Use
- Why Use Animals?
- How to Acquire Animals for Research
- Generality of Animal Research Data
- The Animal Rights Movement
- Animal Research Issues

**Alternatives to Animals in Research: In Vitro Methods and Computer Simulation****Summary****Key Terms**

# Choosing and Using Research Subjects

So far in the research process, you have made several important decisions. You have decided on a topic for your research, have taken an amorphous, broad idea, and honed it into a tight, testable research hypothesis. You also have made some important decisions about the nature of the research design that you will use, the variables you will manipulate and measure, and how you will manipulate and measure those variables. Your next decision involves who will participate in your research study.

A number of important questions must be addressed when choosing subjects for psychological research. Should you use human participants or animal subjects?<sup>1</sup> How will you acquire your sample? What ethical guidelines must you follow when using human participants or animal subjects (see Chapter 7)? If you choose human participants, what is your sample going to look like (age, race, gender, ethnicity, etc.)? If you choose to use animals, where will you get them? What are the implications of choosing one species or strain of a species over another? We explore these and other questions in this chapter. The principles discussed in this chapter apply equally to experimental and nonexperimental research. However, there are additional subject-related issues to consider if your study uses survey methodology. We discuss these issues in Chapter 9, along with other issues concerning survey research methodology.

## GENERAL CONSIDERATIONS

As we have already noted, choosing and using subjects in psychological research requires you to confront several important questions and make several important decisions. The nature of your research may

<sup>1</sup>When discussing those who serve in psychological research, we refer to humans as *participants* and to animals as *subjects*. We also use the term *subjects* when the discussion could apply to either humans or nonhumans (e.g., between-subjects design). The American Psychological Association (APA, 2001) adopted these conventions so we will follow them throughout this book in order to be consistent with APA usage.

drive some of those decisions. For example, if you are experimentally investigating the effects of brain lesions on learning abilities, you must use animal subjects. If you are interested in the dynamics of obedience to authority figures, you must use human participants. However, you may investigate many research areas using either animals or humans (such as operant conditioning, memory, or perception). In these cases, your choice of animals versus humans may depend on the needs of your particular experiment. However, regardless of whether you choose humans or animals, you must consider issues such as ethics, how the subjects will react to your experimental procedure, and the degree of generality of your results.

### Populations and Samples

Imagine that you are interested in investigating the effect of a new computer-based teaching technique on how well eighth graders learn mathematics. Would it be feasible to include *every* eighth grader in the world in your experiment? Obviously not, but what is the alternative? You may have thought to yourself, “I will have to choose *some* eighth graders for the experiment.” If this is what you thought, you are considering an important distinction in research methodology: populations versus samples.

In the hypothetical experiment, you could not hope to include all eighth graders. “All eighth graders” constitutes the **population** under study. Because it is usually not possible to study an entire population, you must be content to study a sample of that population. A **sample** is a small subgroup chosen from the larger population. Figure 6-1 illustrates the relationship between populations and samples.

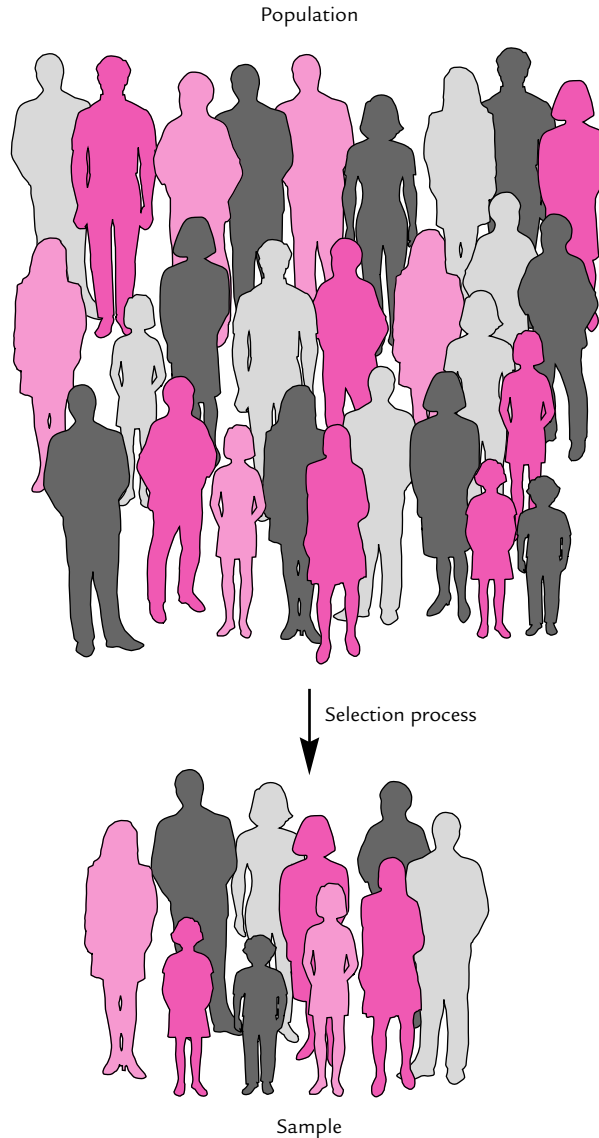
Often researchers find it necessary to define a *subpopulation* for study. In your imaginary study, cost or other factors may limit you to studying a certain region of the country. Your subpopulation might consist of eighth graders from a particular city, town, or district. Furthermore, you might limit yourself to studying certain eighth-grade classes (especially if the school district is too large to allow you to study every class). In this case, you are further dividing your subpopulation. In effect, rather than studying an entire population, you are studying only a small segment of that population.

You can define a population in many ways. For example, if you were interested in how prejudiced attitudes develop in young children, you could define the population as those children enrolled in day-care centers and early elementary school grades. If you were interested in jury decision making, you could define the population as registered voters who are eligible for jury duty. In any case, you may need to limit the nature of the subject population and sample because of special needs of the research.

### QUESTIONS TO PONDER

---

1. How does the nature of your research affect whether you use human participants or animal subjects in your research?
2. What is the difference between a population and a sample?



**FIGURE 6-1** Relationship between population and sample. A sample is a subset of individuals selected from a larger population.

### Sampling and Generalization

An important goal of many research studies is to apply the results, based on a sample of individuals, to the larger population from which the individuals were drawn. You do not want the results from your study of the new teaching techniques to apply only to those eighth graders who participated in the study. Rather, you want your results

to apply to all eighth graders. **Generalization** is the ability to apply findings from a sample to a larger population. In Chapter 4, we noted that studies whose findings can be applied across a variety of research settings and subject populations possess a high degree of *external validity*. Thus, the ability to generalize findings to a larger population contributes to the external validity of a study.

If the results of a study are to generalize to the intended population, you must be careful when you select your sample. The optimal procedure is to identify the population and then draw a **random sample** of individuals from that population. In a random sample, every person in the population has an equal chance of being chosen for the study. (Chapter 9 on using survey research explores the various methods that you can use to acquire a random sample.) A true random sample allows for the highest level of generality from research to real life.

### Nonrandom Sampling

Unfortunately, in psychological research we rarely meet the ideal of having a random sample of individuals from the population. In practice, most psychological studies use a **nonrandom sample**, usually of individuals from a highly specialized subpopulation—college students. In fact, McNemar (1946) characterized psychology as the “science of college sophomores.” Higbee, Millard, and Folkman (1982) report that a majority of studies in social psychology published in the 1970s relied on college students for participants, and there is little to suggest that this practice has changed since then.

We conducted a content analysis of a random sample of articles from the 2006 (volume 90, five articles per issue) volume of the *Journal of Personality and Social Psychology* (the premier journal in social psychology). The analysis showed that 81% of the studies reported in 30 articles used college students exclusively as participants. Another 5.7% used noncollege students (e.g., high school students), and another 5.7% used some combination of student and nonstudent samples. The college student remains the dominant source of research participants, at least in social psychology.

Psychological research uses college students so often because most psychological research is conducted by college professors. For them, college students form a readily available pool of potential research participants. In fact, many psychology departments set up a *subject pool*, usually consisting of introductory psychology students, to provide participants for psychological studies. They are essentially an easily tapped captive pool of individuals. Sampling from a relatively small subject pool is much easier than drawing a random sample from the general population and greatly reduces the time and financial costs of doing research. However, using such a nonrandom sample has a downside. If you use college students in order to save time, effort, and money, you may be sacrificing the generality of your results, and the study will have less external validity. College students differ from the noncollege population in a number of ways (such as in age or socioeconomic status). These differences may limit your ability to apply your results to the larger population beyond college. You may be limited to generalizing only to other college students.

The issue of using college students in research may be overblown (Kardes, 1996). Frank Kardes maintains that college student populations are fine when you are

studying basic psychological processes (e.g., memory) although problems may occur when you are interested in making specific applications of your findings (Kardes, 1996). Research on the issue of student versus nonstudent participants has produced mixed results. A few studies (such as Feild & Barnett, 1978) have found differences between college and noncollege participants. In contrast, Tanford (1984) reports a jury simulation study in which student participants did not differ significantly from “real jurors” on most of the measures included in the study. Given these inconsistent findings, the true impact of using students as participants is difficult to assess. You should recognize that your results may have limited generality.

## QUESTIONS TO PONDER

---

1. What is random sampling and how does it relate to generality of findings?
2. What is nonrandom sampling and what problems does it pose for research?

**Nonrandom Sampling and Internet Research** Studies being conducted on the Internet provide further examples of nonrandom sampling. Participants are self-selected volunteers who participate by filling out Web-page questionnaires or by actively engaging in experimental activities available on the Web. The samples for these studies are composed of individuals who know how to use computers, have access to them, know enough about the Internet to stumble into or otherwise find the studies, and volunteer to participate in them—characteristics that may not be true of many people in the general population. However, proponents of Internet-based research argue that similar problems exist when using traditional subject pools such as those from which college students are drawn. Proponents of Internet research suggest that proper participant recruitment techniques (using postings to various news groups, discussion groups, list servers, and Web sites) are analogous to posting sign-up sheets for a study being offered to members of a traditional subject pool. Proponents argue that proper recruitment techniques actually may lead to a broader range of participants geographically and demographically than do traditional subject pools.

So where do things stand on the Internet sampling issue? John Krantz and Reeshad Dalal (2000) suggest that there are two ways to establish the validity of Web-based research. First, you can compare results from studies (surveys and experiments) done on the Web to results from parallel studies done using more traditional methods. Second, you can evaluate the results from Web-based research to see if they conform to theoretical predictions. Krantz and Dalal conclude that, for the most part, research (survey and experimental studies) conducted via the Internet produces results that are highly similar to those from research done with more conventional methods.

The limited amount of research comparing traditional surveys and Web surveys bears this out (Hamby, Sugarman, & Boney-McCoy, 2006; Huang, 2005; Riva, Teruzzi, & Anolli, 2003). For example, Riva et al. (2003) compared the results from an attitude survey administered via the Internet with those from the same survey administered using a paper-and-pencil format. They found no major differences

between the two methods. They conclude that, given careful attention to sampling, reliability, and validity issues, the Internet can produce results that mirror those from traditional surveys.

A study by Bethell, Fiorillo, Lansky, Hendryx, and Knickman (2004) confirms this. Bethell et al. administered a questionnaire on quality of health care either online or by telephone. The data obtained from the telephone and online surveys were compared to each other and to general population data. Bethell et al. found that the sample for the online survey matched closely the sample of the general population. There was some overrepresentation of respondents in the 45 to 64 age group and some underrepresentation in the 18 to 24 age group. Both the telephone and online surveys underrepresented non-White populations, respondents with less than a high school education, and respondents with annual incomes above \$75,000. Despite these differences, Bethell et al. conclude that the telephone and online samples were representative of the general population. Furthermore, the results obtained from the online survey did not differ significantly from those obtained from existing surveys on the U.S. health-care system.

Although there are striking parallels between the results of Internet and non-Internet studies, this does not mean that all Web-based findings match findings using other methods (Krantz & Dalal, 2000). For example, Michael Link and Ali Mokdad (2005) compared mail, telephone, and Web-survey methods. The survey measured participants' level of alcohol consumption. They found that the Web survey generated a lower response rate (15.4%) than the telephone survey (40.1%) or mail survey (43.6%). They also found that Web-survey respondents in all demographic categories were more likely to report heavy drinking (five or more drinks in a day) than telephone respondents. Link and Mokdad suggest that the higher reported rates of heavy drinking among Web respondents may be due to nonresponse bias.

Clearly, more research is needed on this issue. Overall, the research in this area suggests that the Internet provides a powerful tool for researchers that may have fewer liabilities than critics allege. However, the method may be more problematic when you are asking about sensitive issues such as alcohol consumption.

**Nonrandom Sampling and Animal Subjects** Nonrandom sampling is not restricted to research using human participants. In fact, it is almost standard procedure for research using animal subjects. Laboratory animals are usually ordered for a given study from a single supplier and typically consist of animals of the same species, strain, sex, and age (indeed, many of them may be littermates). A group of 30 female Sprague–Dawley laboratory rats, all 90 days old and obtained from the same supplier, can hardly be considered a random sample of all female laboratory rats, let alone rats in general. In some cases, even supposedly minor differences in strain have been found to alter the results. For example, Helmstetter and Fanselow (1987) showed that the opiate blocker naloxone was effective in suppressing conditioned analgesia (reduced sensitivity to pain) under certain conditions in the Long–Evans strain of laboratory rat but not in the Sprague–Dawley strain. Those who believed that their research on the effects of this drug would generalize from the strain of rat they had selected for testing to laboratory rats in general turned out to be mistaken. However, this problem may be mitigated somewhat if different laboratories attempt to follow up

on initial reports but employ animals of different species, strain, sex, or age, or even similar animals from a different supplier. If the original findings are restricted to a given species, strain, sex, age, or supplier, they will fail to replicate in laboratories in which these factors differ. In fact, Helmstetter and Fanselow's study was prompted by such a replication failure.

### Is Random Sampling Always Necessary?

The highest level of generality will flow from research using a true random sample. However, is it necessary for all kinds of research to have high levels of generality (external validity)? As we noted in Chapter 4, perhaps not. Random sampling is especially necessary when you want to apply your findings directly to a population (Mook, 1983; Stanovich, 1986). Political polls, for example, have such a requirement. Pollsters want to predict accurately the percentage of voters who will vote for candidate A or B. That is, pollsters try to predict a specific behavior (e.g., voting) in a specific set of circumstances (Stanovich, 1986). Mook (1983), however, suggests that most research in psychology does not have a specific-to-specific application. In fact, the goal of most psychological research is to predict from the general level (e.g., a theory) to the specific (specific behavior; Stanovich, 1986). Most findings from psychological research are applied *indirectly* through theories and models (Stanovich, 1986). According to Stanovich, many applications of psychological research operate indirectly through the impact of their theories, thus making random samples less of a concern.

Factors other than the nature of the sample that affect the generality of your results include the realism of the research setting and the way in which you manipulate the independent variables. Other sampling considerations most relevant to non-experimental research are discussed in Chapters 8 and 9.

### QUESTIONS TO PONDER

---

1. How does nonrandom sampling apply to Internet research?
2. What does research tell us about sampling issues relating to Internet research?
3. How does nonrandom sampling apply to animal research?
4. In what types of research might random sampling be less important?

### ACQUIRING HUMAN PARTICIPANTS FOR RESEARCH

---

Whether your research is experimental or nonexperimental, you must consider three factors when acquiring participants for your research: You must consider (1) the setting in which your research will take place, (2) any special needs of your particular research, and (3) any institutional, departmental, and ethical policies and guidelines governing the use of participants in research.



## The Research Setting

Chapter 4 distinguished between laboratory and field research. In field research, you conduct your research in the participant's natural environment, whereas in laboratory research you bring your participants into a laboratory environment of your creation. Acquiring participants for laboratory research differs from acquiring them for field research.

**Laboratory Research** If you choose to conduct your research in a laboratory setting, there are two principal ways of acquiring participants. First, you can solicit volunteers from whatever participant population is available. For example, you could recruit participants from your university library or lounge area. These participants would participate on a voluntary basis. As we indicate later in this chapter, voluntary participation has both positive and negative consequences for your research. Second, you can use a subject pool if one exists. Individuals in the subject pool may be required to participate in a certain number of studies (with an alternative to the research option provided). If you adopt this strategy, you must make sure that your recruitment procedures do not coerce folks into participating. Even when using a subject pool, participation in a research study must be voluntary.

**Field Research** Field research requires you to select your participants while they are in their natural environment. How you acquire your participants for field research depends on the nature of your study. For example, if you are conducting a survey, you would use one of the survey sampling techniques discussed in Chapter 9 to acquire a sample of participants. Essentially, these techniques involve selecting a participant from a population, contacting that person, and having him or her complete your questionnaire.

If you were running a field experiment, you could use one of two methods for acquiring participants, again depending on the nature and needs of your study. Some field experiments are actually carried out much like laboratory experiments except that you take your “show” (equipment, assistants, measuring devices, etc.) on the road and set up in the participant's natural environment. This is what Sheina Orbell and Martin Hagger (2006) did in a field experiment investigating how adults respond to persuasive messages about the effects of diabetes.

Participants were recruited by having researchers do home visits in a particular town. Participants were invited to take part in a study of their attitudes about taking part in a diabetes screening program and were asked to complete a questionnaire about participation in such a program. In this experiment, participants were randomly assigned to one of two versions of a persuasive appeal. One paragraph of the instructions for the questionnaire pointed out the positive and negative consequences of participating in the screening program. The main independent variable was the “time frame” for the positive and negative aspects of participation. In one condition, the positive aspects were said to be long term (participating in screening gives people peace of mind for years to come) and the negative consequences short term (undergoing unpleasant procedures immediately). In the other condition, the positive consequences were cast as short term (“getting immediate peace of mind” about their health) and the negative consequences in the long term (worrying about taking medicine for their whole lives).

In this type of field experiment, the researchers maintain about as much control over participant selection and assignment as they would if the experiment were conducted in the laboratory. However, the researchers are at the mercy of whoever happens to be at home on any given day. Thus, with field research, you have less control over participants than in the laboratory.

In another type of field experiment, you set up a situation and wait for participants to happen along. A field experiment reported by Michael Shohat and Jochen Musch (2003) conducted in Germany illustrates this strategy. Shohat and Musch were interested in studying ethnic discrimination. They set up auctions on eBay to sell DVDs, and manipulated the ethnicity of the seller. On one eBay listing, the seller had a Turkish username. On another, the seller had a German username. The researchers recorded the number of hits on each listing as well as the average price paid for the DVDs. In this kind of field experiment, you have less control over who participates in your research. Whoever happens to sign in to eBay at a particular time and search for DVDs would be potential participants.

### The Needs of Your Research

Special needs of your research may affect how you acquire participants. In some cases, you may have to screen potential participants for certain characteristics (such as gender, age, or personality characteristics). For example, in a jury study, you might screen participants for their level of authoritarianism and include only authoritarians in your research. To do this, you must first pretest participants using some measure of authoritarianism and then recruit only those who fall into the category you want. Bear in mind that doing this affects the external validity of your findings. The results you obtain with participants who score high in authoritarianism may not apply to those who show lower levels of authoritarianism.

As another example, you may need children of certain ages for a developmental study of intelligence. Acquiring a sample of children for your research is a bit more involved than acquiring a sample of adults. You must obtain permission from the child's parent or guardian, as well as from the child him- or herself. In practice, some parents may not want their children to participate. This again raises issues of external validity. Your sample of children of parents who agree to allow participation may differ from the general population of children.

### Institutional Policies and Ethical Guidelines

All psychological research involving human participants must comply with the ethical guidelines set out by the American Psychological Association (APA) and with federal and state laws regulating such research. (We discuss these requirements in the next chapter.) Institutions have their own rules concerning how human participants can be recruited and used in research. Although these rules must conform to relevant ethical codes and laws, there is considerable latitude when it comes to setting up subject pools. During the planning stages of your research, you should familiarize yourself with the federal and state laws concerning research using human participants, as well as the policies of the institution in which you are conducting your research.

## QUESTIONS TO PONDER

---

1. How does the setting for your research affect participant recruitment?
2. How do the needs of your research influence participant recruitment?
3. How do institutional policies and ethical considerations affect participant recruitment?

## VOLUNTARY PARTICIPATION AND VALIDITY

---

Participants must voluntarily agree to be in your research. This raises an important question: Are volunteer participants representative of the general population? Individuals who choose to participate in research undoubtedly differ those who do not. Because a sample made up entirely of volunteers is biased, the external validity of your experiment may be affected; this is known as the **volunteer bias**.

There are two assumptions inherent in the previous discussion: (1) Volunteers differ in meaningful ways from nonvolunteers, and (2) the differences between volunteers and nonvolunteers affect the external validity of your research.

### Factors That Affect the Decision to Volunteer

Two categories of factors affect a prospective participant's decision to volunteer: characteristics of the participant and situational factors. We explore each of these next.

**Participant-Related Characteristics** Rosenthal and Rosnow (1975) provide the most comprehensive study of the characteristics of the volunteer subject in their book *The Volunteer Subject*. Table 6-1 lists several characteristics that, according to Rosenthal and Rosnow, distinguish volunteers from nonvolunteers. Associated with each characteristic is the degree of confidence Rosenthal and Rosnow believe you can have in the validity of each attribute.

Whether a person volunteers for a study and how that person performs may depend on a combination of personal and study characteristics. For example, Rosenthal and Rosnow (1975) point out that firstborns may respond more frequently than later-borns to an "intimate" recruitment style for an experiment dealing with group dynamics. Later-borns may respond more frequently than firstborns to a request for participants in an experiment involving stress. Similarly, a sociable person may be more likely to volunteer for an experiment that is "sociable" in nature and less likely to volunteer for an experiment in which there is little or no contact with others. Also, volunteers may show better adjustment than nonvolunteers in experiments that require self-disclosure. Other research suggests that volunteers also may be more field dependent (rely heavily on environmental cues) than nonvolunteers (Cooperman, 1980) and more willing to endure higher levels of stress in an experiment (Saunders, 1980).

A more recent study by Bernd Marcus and Astrid Schütz (2005) sought to relate dimensions of the "big-five" personality dimensions (agreeableness, openness to new experience, conscientiousness, extroversion/introversion, and neuroticism)

**TABLE 6-1 Characteristics of People Who Volunteer for Research****MAXIMUM CONFIDENCE**

1. Volunteers tend to be more highly educated than nonvolunteers.
2. Volunteers tend to come from a higher social class than nonvolunteers.
3. Volunteers are of higher intelligence in general, but not when volunteering for atypical research (such as hypnosis, sex research).
4. Volunteers have a higher need for approval than nonvolunteers.
5. Volunteers are more social than nonvolunteers.

**CONSIDERABLE CONFIDENCE**

1. Volunteers are more “arousal seeking” than nonvolunteers (especially when the research involves stress).
2. People who volunteer for sex research are more unconventional than nonvolunteers.
3. Females are more likely to volunteer than males, except where the research involves physical or emotional stress.
4. Volunteers are less authoritarian than nonvolunteers.
5. Jews are more likely to volunteer than Protestants; however, Protestants are more likely to volunteer than Catholics.
6. Volunteers have a tendency to be less conforming than nonvolunteers, except where the volunteers are female and the research is clinically oriented.

SOURCE: Adapted from Rosenthal and Rosnow, 1975.

to willingness to respond to a survey. Marcus and Schütz identified several personal Web sites (i.e., Web sites where people post personal information about themselves). A group of observers evaluated the Web sites and characterized the personality of each of the people who maintained the sites. This observer evaluation provided the measures of the big-five personality dimensions. Marcus and Schütz then contacted the owners of the Web sites, asked them to participate in a survey on the psychology of personal Web sites, and recorded the extent to which they completed the survey. This provided the measure of whether Web-site owners were willing to respond to a survey. Marcus and Schütz found that those who responded to the survey were rated as more agreeable and open to new experience than those who did not respond. However, responders and nonresponders did not differ on the conscientiousness dimension. Omission of items on the survey was significantly related to low levels of openness to new experience. Marcus and Schütz suggest that their findings have implications for the validity of research where personality profiles of participants are compared to normative data. They point out, for example, that those who volunteer to complete surveys differ from those who do not. The results from such studies may not generalize well to the general population. In another study, researchers

found a difference between volunteers and nonvolunteers on the dimensions of conscientiousness and neuroticism (Lönqvist, Paunonen, Verkaslo, Leikas, Tuulio-Henriksson, & Lönqvist, 2006). In this study volunteers (compared to nonvolunteers) were lower in neuroticism and higher in conscientiousness.

One area where volunteer bias might be a particular problem is research on sexual functioning. Based on a review of the literature, Boynton (2003) concluded that women were less likely to volunteer for research on sexual behavior than men and were more likely than men to refuse to answer certain questions about sexuality. Generally, individuals who are comfortable with sexuality are more likely to volunteer for this type of research than those who are less comfortable (Boynton, 2003). A similar finding was obtained by Nirenberg et al. (1991), who compared alcoholics who volunteered to participate in a study on sexual functioning and behavior with alcoholics who declined to participate. Volunteers expressed greater interest in sex, less satisfaction with sex, higher rates of premature ejaculation, and more concern over sexual functioning than nonvolunteers. Additionally, volunteers used substance-abuse counseling more often and had higher rates of drug use. On the other hand, Mandel, Weiner, Kaplan, Pelcovitz, and Labruna (2000) found few differences between volunteer and nonvolunteer samples of abused families. In fact, Mandel et al. report that there were far more similarities than dissimilarities between the volunteers and nonvolunteers.

Where does all of this leave us? It is clear that under some circumstances volunteers and nonvolunteers differ. These differences may translate into lower external validity for findings based on volunteer participant samples. The best advice we can give is to be aware of the potential for volunteer bias and take it into account when interpreting your results.

**Situational Factors** In addition to participant characteristics, situational factors may affect a person's decision to volunteer for behavioral research. According to Rosenthal and Rosnow (1975), you can have "maximum confidence" in the conclusion that people who are more interested in the topic being researched and who have expectations of being favorably evaluated will be more likely to volunteer for a particular research study. You can have "considerable confidence" that if potential participants perceive the research as being important, feel guilty about not participating, and are offered incentives to participate, they will be more likely to volunteer. Other factors that have less impact on the decision include personal characteristics of the person recruiting the participants, the amount of stress inherent in the experiment, and the degree to which participants feel that volunteering is the "normative, expected, appropriate thing to do" (Rosenthal & Rosnow, 1975, p. 119). Finally, you can have only "minimum confidence" that a personal acquaintance with the recruiter or public commitment to volunteering will affect the rate of volunteering.

As with the participant-related factors, the operation of the situational factors may be complex. For example, people are generally less disposed to volunteer for experiments that involve stress or aversive situations. According to Rosenthal and Rosnow (1975), the personal characteristics of the potential participant and the nature of the incentives offered may mediate the decision to volunteer for this type of research. Also, stable personal characteristics may mediate the impact of offering material rewards for participation in research.

The general conclusion from the research of Rosenthal and Rosnow (1975) is that several participant-related and situational characteristics affect an individual's decision about volunteering for a particular research study. Such a decision may be influenced by a variety of factors that interact with one another. In any case, it is apparent that volunteering is not a simple random process. Certain types of people are disposed to volunteer generally and others for certain specific types of research.

The nature of the stimuli used in a study also affects the likelihood of volunteering. For example, Gaither, Sellbom and Meier (2003) had men and women fill out a questionnaire asking them whether they would be willing to participate in research in which a variety of different sexually explicit images were to be judged. Gaither et al. found that men were more likely than women to volunteer for research involving viewing images of heterosexual sexual behavior, nude women, and female homosexual sexual behavior. Women were more likely than men to volunteer for research involving viewing images of nude men and male homosexual sexual behavior. Gaither et al. also found that regardless of gender, those willing to volunteer for this type of research were higher in sexual and nonsexual sensation seeking.

Finally, media coverage may relate to willingness to volunteer. Gary Mans and Christopher Stream (2006) investigated the relationship between the amount and nature of media coverage and volunteering for medical research. Mans and Stream found that the greater the *positive* media coverage of a study, the more willing people are to volunteer (although the converse was not true). There was no relationship between the *amount* of media coverage and willingness to volunteer, however. So, something beyond your control, like media coverage, can affect a person's willingness to volunteer for your research.

## QUESTIONS TO PONDER

---

1. What is the volunteer bias and why is it important to consider in your research?
2. How do volunteers and nonvolunteers differ in terms of personality and other characteristics?
3. What are some of the situational factors that affect a participant's decision to volunteer?

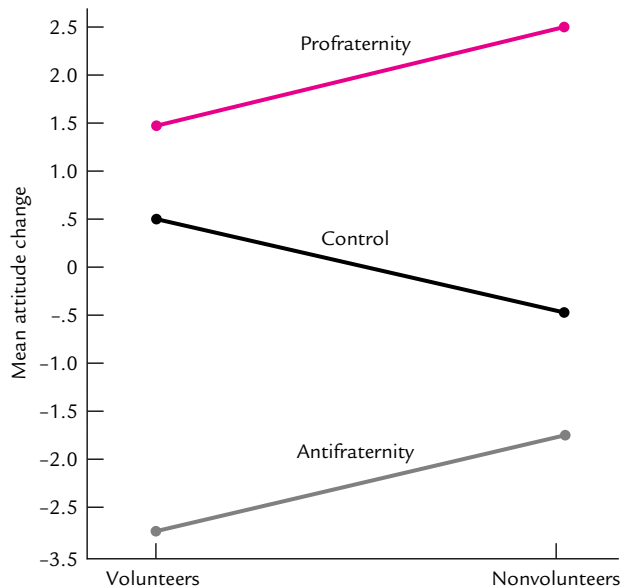
## Volunteerism and Internal Validity

Ideally, you want to establish that variation in your independent variable *causes* observed variation in your dependent variable. However, variables related to voluntary participation may, quite subtly, cause variation in your dependent variable. If you conclude that the variation in your independent variable caused the observed effects, you may be mistaken. Thus, volunteerism may affect "inferred causality" (Rosenthal & Rosnow, 1975), which closely relates to internal validity.

Rosenthal and Rosnow (1975) conducted a series of experiments investigating the impact of volunteering on inferred causality within the context of an attitude

change experiment. In the first experiment, 42 undergraduate women (20 of whom had previously indicated their willingness to volunteer for a study) were given an attitude questionnaire concerning fraternities on college campuses. A week later, the experimenters randomly assigned some participants to a profraternity communication, others to an antifraternity communication, and still others to no persuasive communication. The participants were then given a measure of their attitudes toward fraternities.

Although the persuasive communication changed attitudes more than the other types, the volunteers were more affected by the antifraternity communication than were nonvolunteers, as shown in Figure 6-2. A tentative explanation offered by Rosenthal and Rosnow (1975) for this effect centered on the higher need for approval among volunteers than nonvolunteers. Volunteers tended to see the experimenter as being antifraternity (although only slightly). Apparently, the volunteers were more motivated to please the experimenter than were the nonvolunteers. The desire to please the experimenter, not the content of the persuasive measure, may have caused the observed attitude change. The results of this experiment show that variables relating to voluntary participation may cloud any causal inferences you draw about the relationship between your independent and dependent variables. Rosenthal and Rosnow conclude that “subjects’ reactions to a persuasive communication can be largely predicted from their original willingness to participate in the research” (1975, p. 155). According to Rosenthal and Rosnow, the volunteers’ predisposition to comply with demand characteristics of the experiment indicates that volunteerism serves as a “motivation mediator” and may affect the internal validity of an experiment.



**FIGURE 6-2** Attitude change as a function of type of message and volunteerism.

SOURCE: Based on the data from Rosenthal and Rosnow, 1975.

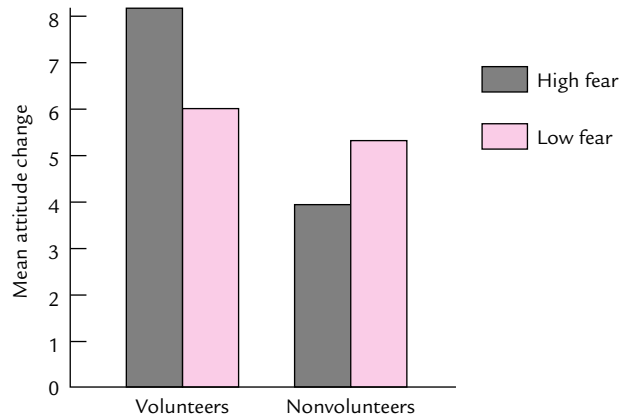
### Volunteerism and External Validity

Ideally we would like the results of our research to generalize beyond our sample. Volunteerism may affect our ability to generalize, thus reducing external validity. If volunteer participants have unique characteristics, your findings may apply only to participants with those characteristics.

There is evidence that individuals who volunteer for certain types of research differ from nonvolunteers. For example, Davis et al. (1999) found that individuals high on a measure of empathy were more likely to volunteer for a sympathy-arousing activity than those lower in empathy. In another study, Carnahan and McFarland (2007) investigated whether individuals who volunteered for a “study of prison life” (such as the Stanford Prison Study described in detail later in this chapter) differed from those who volunteered for an identically described study omitting the reference to “prison life.” Carnahan and McFarland found that individuals who volunteered for the prison life study were higher on aggressiveness, right wing authoritarianism (a measure of submissiveness to authority), Machiavellianism (a tendency to mistrust others), narcissism (a need for power and negative responses to threats to self-esteem), and social dominance (the desire for one’s group to dominate others) than those who volunteered for a “psychological study.” Additionally, those who volunteered for the “psychological study” were higher on empathy and altruism. In another study, women who were willing to volunteer for a study of sexual arousal to viewing erotic materials using a vaginal measure of arousal were more likely to have experienced sexual trauma and had fewer objections to viewing erotic material than nonvolunteers (Wolchik, Spencer, & Lisi, 1983). Finally, volunteers and nonvolunteers react differently to persuasive communications using fear (Horowitz, 1969). As shown in Figure 6-3, volunteers showed more attitude change in response to a high-fear communication than did the nonvolunteers. However, little difference emerged between volunteers and nonvolunteers in a low-fear condition.

The results of these studies suggest that using volunteer participants may yield results that do not generalize to the general population. For example, the results from Carnahan and McFarland’s (2007) study suggest that how participants in the original

**FIGURE 6-3** Attitude change as a function of fear arousal and volunteerism.  
SOURCE: Based on an experiment by Horowitz, 1969.





Stanford Prison study (e.g., participants randomly assigned to be guards acting cruelly) responded may not represent how people in general would respond in such a situation. The reaction observed in that study may be limited to those who are predisposed to react cruelly. Similarly, findings relating to how women respond to erotica may not apply to all women. In both examples the special characteristics of the volunteers limits the generality of the findings.

### Remedies for Volunteerism

Are there any remedies for the “volunteerism” problem? Rosenthal and Rosnow (1975, pp. 198–199) list the following actions you can take to reduce the bias inherent in the recruitment of volunteers:

1. Make the appeal for participants as interesting as possible, keeping in mind the nature of the target population.
2. Make the appeal as nonthreatening as possible so that potential volunteers will not be “put off” by unwarranted fears of unfavorable evaluation.
3. Explicitly state the theoretical and practical importance of the research for which volunteering is requested.
4. Explicitly state in what way the target population is particularly relevant to the research being conducted and the responsibility of the potential volunteers to participate in research that has the potential for benefiting others.
5. When possible, potential volunteers should be offered not only pay for participation but also small courtesy gifts simply for taking the time to consider whether they will want to participate.
6. Have the request for volunteering made by a person of status as high as possible and preferably by a woman.
7. When possible, avoid research tasks that may be psychologically or biologically stressful.
8. When possible, communicate the normative nature of the volunteering response.
9. After a target population has been defined, make an effort to have someone known to that population make an appeal for volunteers. The request for volunteers itself may be more successful if a personalized appeal is made.
10. For situations in which volunteering is regarded by the target population as normative, conditions of public commitment to volunteer may be more successful. If nonvolunteering is regarded as normative, conditions of private commitment may be more successful.

### QUESTIONS TO PONDER

---

1. How does the volunteer bias relate to internal and external validity?
2. What are some of the remedies for the problem of volunteer bias?

## RESEARCH USING DECEPTION

---

Imagine that you are riding a bus home from school. All of a sudden someone from the back of the bus staggers past you and falls down, hitting his head. You notice a small trickle of blood running down the side of the victim's mouth. You are both alarmed and concerned about the victim, but you don't get up to help. However, you see several others are going to the victim's aid. At that point, a person at the front of the bus stands up and informs you that you were all participants in a psychological experiment on helping behavior. You have just been a participant in an experiment using deception. How would you feel about this situation? Would you be relieved that the "victim" was not really hurt? Would you be angry that the researchers included you in their experiment without your knowing about it?

In most cases, psychological research involves fully informing participants of the purposes and nature of an experiment. Participants in research on basic processes such as perception and memory are usually informed beforehand of what the experiment will involve. However, in some cases, full disclosure of the nature and purpose of your study would invalidate your research findings. When you either actively mislead participants or purposely withhold information from the participant, you are using **deception**. Although the use of deception in research declined (Sieber, Iannuzzo, & Rodriguez, 1995) between 1969 and 1995, it is still used for some research applications. However, using deception is very controversial, with opponents and proponents on both sides of the issue (Pittinger, 2002).

Why use deception? There are two main reasons (Hertwig & Ortmann, 2008). First, deception allows you to create interesting situations that are not likely to occur naturally and then study the reactions of individuals who experience them (Hertwig & Ortmann, 2008). This is what you experienced in our hypothetical helping study previously mentioned. It is much more efficient to create an emergency situation (a person falling down on a bus) than it is to wait around for one to occur on its own. Second, there are certain aspects of behavior that can only be studied if a person is caught off guard (Hertwig & Ortmann, 2008). Again, this second factor was evident in the example that opened this section. Proponents of deception argue that if you told people beforehand that an actor was going to fall down in front of them, they would behave differently than if the same situation were presented without the prior information.

In the sections that follow, we discuss how deception is used, its effects on research participants, and possible remedies to the problems inherent in using deception in research.

### Types of Research Deception

Deception may take a variety of forms. Arellano-Galdamas (1972, cited in Schuler, 1982) distinguishes between active and passive deception. *Active deception* includes the following behavior (Schuler, 1982, p. 79):

1. Misrepresenting the purposes of the research
2. Making false statements as to the identity of the researcher
3. Making false promises to the participants

4. Violating a promise to keep the participant anonymous
5. Providing misleading explanations of equipment and procedures
6. Using pseudosubjects (people who pose as participants but work for the experimenter)
7. Making false diagnoses and other reports
8. Using false interaction
9. Using placebos and secret administration of drugs
10. Providing misleading settings for the investigations and corresponding behavior by the experimenter

*Passive deception* includes the following (Schuler, 1982, p. 79):

1. Unrecognized conditioning
2. Provocation and secret recording of negatively evaluated behavior
3. Concealed observation
4. Unrecognized participant observation
5. Use of projective techniques and other personality tests

### Problems Involved in Using Deception

Although deception is a popular research tactic (especially in social psychology), some researchers consider it inappropriate (Kelman, 1967). In fact, deception does pose a number of problems for both the participant and the experimenter. For example, research suggests that once deceived, participants may react differently from nondeceived participants in a subsequent experiment (Silverman, Shulman, & Weisenthal, 1970).

Deception may also influence whether a person would be willing to volunteer for future research. Generally, research participants have a negative view of deception and indicate that they would be less likely to participate in a subsequent study if they were deceived in an earlier one (Blatchley & O'Brien, 2007). Participants in Blatchley and O'Brien's study also indicated that the more frequently deception was seen as part of psychological research, the less likely they would be to participate in a subsequent study. Blatchley and O'Brien concluded that frequent use of deception in research results in negative attitudes toward research and toward psychology as a whole, resulting in a "reputational spillover effect" (Blatchley & O'Brien, 2007, p. 527).

Despite the evidence for a reputational spillover effect, the news about using deception is not all bad. There are situations in which research participants understand the need for deception. For example, Aguinis and Henle (2001) investigated potential research participants' reactions to a technique called the "bogus pipeline." This technique assesses attitudes by hooking up participants to a machine they believe will tell a researcher whether they are telling the truth about their attitudes. The catch is that the machine does nothing. However, research has found that the bogus pipeline procedure elicits more accurate measures of attitudes than conventional questionnaires. Participants in Aguinis and Henle's study are given a summary of a

published article that used the bogus pipeline to assess attitudes and are then asked several questions about the technique. Aguinis and Henle found that even though their respondents believed that the participants in the study would react negatively to the deception involved in the bogus pipeline, they saw the technique as a valuable tool for getting truthful information and believed that the benefits derived from the technique outweighed the costs.

Another problem with deception is that the participant in a deception experiment has been exposed to an elaborate hoax. Most research participants are not expecting to be deceived in an experiment (Blatchley & O'Brien, 2007). Being deceived in an experiment may violate an assumed trust between the participant and the researcher. As a consequence, a participant who has been deceived may feel betrayed and duped by the experimenter. The participant may experience a loss of self-esteem or develop a negative attitude toward research. According to Holmes (1976a), the researcher's responsibility is to "dehoax" the participant after the experiment.

Yet another problem may arise from deception research if, during the course of the experiment, participants find out something disturbing about themselves. Holmes (1976b) maintains that the researcher has the responsibility to desensitize participants concerning their own behaviors.

Stanley Milgram's (1963, 1974) classic obedience research illustrates these problems of deception research. Briefly, Milgram led participants to believe they were participating in an experiment investigating the effects of punishment on learning. The participant was told to deliver an electric shock to the "learner" each time the learner made an error. The shock intensity was to be increased following each delivery of shock. In reality, the assignment of individuals to the role of "teacher" or "learner" was prearranged (the "learner" was a confederate of the researcher), and no real shocks were being delivered by the participants. The true purpose of the research was to test the participant's obedience to an authority figure (the experimenter) who insisted that the participant continue with the procedure whenever the participant protested.

Milgram's research relied heavily on deception. Participants were deceived into believing that the experiment was about learning and that they were actually delivering painful electric shocks to another person. The problem of hoaxing is evident in this experiment. Participants also were induced to behave in a highly unacceptable manner. The participant may have found out that he or she was "the type of person who would intentionally hurt another," an obvious threat to the participant's self-concept. To be fair, Milgram did extensively debrief his participants to help reduce the impact of the experimental manipulations. (Debriefing is discussed later in the chapter.) However, some social psychologists (e.g., Baumrind, 1964) maintain that the experiment was still unethical.

Ethical treatment of participants requires that you inform participants of the nature and purpose of the research before they participate. Does deception on the part of the researcher constitute unethical behavior? According to the APA ethical principles (2002), deception may be used only if the experimenter can justify the use of deception based on the study's scientific, educational, or applied value; if alternative procedures that do not use deception are not available; and if the participants are provided with an explanation for the deception as soon as possible. The APA code of ethics thus allows researchers to use deception but only under restricted conditions.

## QUESTIONS TO PONDER

---

1. What is deception and why is it used in research?
2. What are the different types of research deception?
3. What are the problems created by using deception?

### Solutions to the Problem of Deception

Obviously, deception may result in ethical and practical problems for your research. To avoid these problems, researchers have suggested some solutions. These range from eliminating deception completely and substituting an alternative method called *role playing* to retaining deception but adopting methods to soften its impact.

**Role Playing** As an alternative to deception, critics have suggested using **role playing**. In role playing, participants are fully informed about the nature of the research and are asked to act as though they were subjected to a particular treatment condition. The technique thus relies on a participant's ability to assume and play out a given role.

Some studies demonstrate that participants *can* become immersed in a role and act accordingly. The famous Stanford prison study is one such example. In that study, participants were randomly assigned to the role of either prisoners or guards in a simulated prison. Observations were made of the interactions between the "guards" and their "prisoners" (Haney, Banks, & Zimbardo, 1973). Participants were able to play out their roles even though they were fully aware of the experimental nature of the situation. Similarly, Janis and Mann (1965) directly tested the impact of emotional role playing by having participants assume the role of a dying cancer patient. Other, non-role-playing participants did not assume the role but were exposed to the same information as the role-playing participants. Participants in the role-playing condition showed more attitudinal and behavioral changes than participants in the non-role-playing control group.

Participants are thus *capable* of assuming a role. The next question is whether the data obtained from role-playing participants are equivalent to the data generated from deception methods.

Opponents of role playing have likened the practice of role playing to "the days of prescientific techniques when intuition and consensus took the place of data" (Freedman, 1969, p. 100). They contend that participants fully informed of the nature and purposes of research will produce results qualitatively different from those produced from uninformed participants.

Resnick and Schwartz (1973) provide support for this view. In a simple verbal conditioning experiment (statements that use "I" or "we" were reinforced), some participants were fully informed of the reinforcement manipulation whereas others were not. The results showed that the uninformed participants displayed the usual learning curve (using more "I-we" statements in the reinforcement condition). In contrast, the fully informed volunteers showed a decline in the rate of "I-we" statements. Thus, informed and uninformed participants behaved differently. Other

research (Horowitz & Rothschild, 1970) has provided additional evidence that role-playing techniques are not equivalent to deception methods.

The use of deception raises questions of both ethics and sound methodology. Role playing has not been the panacea for the problems of deception research. For this reason, deception continues to be used in psychological research (most often in social psychological research). Given that you may decide to use deception in your research, are there any steps that you can take to deal with the ethical questions about deception and reduce the impact of deception on participants? The answer to this question is a qualified yes.

**Obtaining Prior Consent to Be Deceived** Campbell (1969) suggests that participants in a subject pool be told at the beginning of the semester that some experiments in which they may participate might involve deception. They could be provided with an explanation of the need for deception at that time. Gamson, Fireman, and Rytina (1982) devised an additional ingenious method for the securing of informed consent to be deceived. Participants were contacted and asked to indicate the types of research in which they would be willing to participate. Included in the list was research in which the participants were not fully informed. With this strategy, you might choose only those participants who agree to be deceived. Of course, choosing only the agreeable participants may contribute to sampling error and affect external validity.

**Debriefing** Even if you can quell your conscience about the ethical aspects of deception by obtaining prior consent to deceive, you are still obligated to your participants to inform them of the deception as soon as possible after the research. A technique commonly used to do this is **debriefing**.

During a debriefing session, you inform your participants about the nature of the deception used and why the deception was necessary. Because knowledge of having been deceived may lead to bad feelings on the part of the participant, one goal of the debriefing session should be to restore the participant's trust and self-esteem. You want the participant to leave the experiment feeling good about the research experience and less suspicious of other research.

Research shows that debriefing has become more frequent in research (Ullman & Jackson, 1982). Ullman and Jackson showed that only 12% of studies published in two major social psychology journals reported using debriefing in 1964. In contrast, 47% were found to have used debriefing in 1980. Clearly, researchers are becoming sensitive to the problems of deception research and have begun to use debriefing more.

But is debriefing effective? Research on this issue has yielded conflicting results. Walster, Berscheid, Abrahams, and Aronson (1967) found that the effects of deception persisted even after debriefing. In contrast, Smith and Richardson (1983) report that debriefing was successful in removing negative feelings about deception research. They conclude that effective debriefing not only can reverse the ill effects of deception but also can help make participants who felt harmed by research become more positive about the research experience.

In an experiment by Nicholas Epley and Chuck Huff (1998), participants served in a replication of a deception experiment. During the first experimental session, participants completed several tasks, including completion of a self-efficacy scale and a

task requiring them to read short essays and answer questions about them. Half of the participants were given positive feedback about their performance on the essay task, and half were given negative feedback. At the end of the first session, half of the participants received full debriefing that explained the deception (false feedback). The remaining participants were partially debriefed, not including a description of the deception. In two subsequent sessions, participants completed several measures, some of which were the same measures completed in the first session.

Epley and Huff (1998) found that participants generally reported positive reactions to being in the experiment, regardless of whether they received full or partial debriefing. However, participants who were fully debriefed indicated greater suspicion concerning future experiments than those who were only partially debriefed. The suspicion over future research persisted and actually gained strength over three months. Generally, participants did not have strong negative reactions to being in a deception experiment. Apparently, deception is not as costly and negative to research participants as previously believed (Epley & Huff, 1998).

A possible resolution to this conflict in results emerges from an evaluation of different debriefing techniques. Smith and Richardson (1983) point out that “effective” debriefing can reverse the negative feelings associated with deception. But what constitutes “effective” debriefing?

An answer to this question can be found in a study reported by Ross, Lepper, and Hubbard (1975). This study found that the effects of false feedback about task performance persevered beyond debriefing. When participants were presented with “outcome” debriefing (which merely pointed out the deception and justified it), the effects of deception persevered. In contrast, if participants were told that sometimes the effects of experimental manipulations persist after the experiment is over, the debriefing was more successful.

There is another component you can add to standard outcome debriefing that can increase its effectiveness (McFarland, Cheam, & Buehler, 2007). Cathy McFarland, Adeline Cheam, and Roger Buehler report that in addition to informing participants that the test results provided by experimenters are false, participants should be informed that the test itself was bogus. When participants were told of the bogus nature of the test during debriefing, perseverance effects were reduced markedly. Oczak and Niedźwieńska (2007) tested the effectiveness of an even more extensive debriefing procedure. In the expanded debriefing, the mechanisms used in the debriefing were explained and participants were given practice detecting and countering deception. The extended debriefing procedure, according to Oczak and Niedźwieńska, allows participants to effectively recognize and cope with future attempts to deceive them. When the expanded debriefing was compared to a standard debriefing procedure, Oczak and Niedźwieńska found that participants exposed to the expanded procedure reported a more positive mood and a more positive attitude toward research than those exposed to the standard procedure. These two studies suggest that expanding debriefing to address deception more effectively can remove negative effects of deception and help counter negative attitudes toward research.

Although no easy answers to the problems generated by using deception can be found, some insight on how to soften the effects of deceptive strategies might help solve the problems. First, carefully consider the ethical implications of deception

before using it. You, the researcher, are ultimately responsible for treating your participants ethically. If deception is necessary, you should take steps both to dehoax and to desensitize participants through debriefing (Holmes, 1976a, 1976b). The debriefing session should be conducted as soon as possible after the experimental manipulations and should include

1. A full disclosure of the purposes of the experiment.
2. A complete description of the deception used and a thorough explanation of why the deception was necessary.
3. A discussion of the problem of perseverance of the effects of the experimental manipulations.
4. A convincing argument for the necessity of the deception. You also should convince the participant that the research is scientifically important and has potential applications.

During debriefing, be as sincere with the participants as possible. The participant has already been “duped” in your experiment. The last thing that the participant needs is an experimenter who behaves in a condescending manner during debriefing (Aronson & Carlsmith, 1968). Despite the deception used, make the participant recognize that he or she was an important part of the research.

One final question about debriefing: Will the participant believe your debriefing? That is, will the person who has already been deceived believe the experimenter’s assertions made during debriefing? Holmes (1976a) points out that there is no guarantee that the participants will believe the experimenter during debriefing. According to Holmes, participants may feel that they are being set up for another deception. The researcher may need to take some drastic measures to ensure that the participant leaves the experiment believing the debriefing. Holmes (1976a) suggests the following options:

1. Use demonstrations for the participant. For example, the participant could be shown that the experimenter never saw the participant’s actual responses (this would be effective when false feedback is given) or that the equipment used to monitor the participant was bogus.
2. Allow the participants to observe a subsequent experimental session showing another participant receiving the deception.
3. Give participants an active role in the research. For example, the participant could serve as a confederate in a subsequent experimental session.

Complete and honest debriefing is designed to make the participant feel more comfortable about deceptive research practices. Whereas this goal may be accomplished to some degree, the integrity of your research may be compromised. If your participants tell other prospective participants about your experiment (especially in cases in which deception is used), subsequent data may be invalid. Consequently, it’s a good idea to ask participants not to discuss with anyone else the nature of your experiment. Point out to the participants that any disclosure of the deception or any other information about your experiment will invalidate your results. Your goal



should be to have your participant understand and agree that not disclosing information about your experiment is important.

Debriefing is not used exclusively for research using deception. In fact, it is good, ethical research practice to debrief participants after *any* experiment. During such a debriefing session, the participants should be given a full explanation of the methods used in the experiment, the purpose of the experiment, and any results available. Of course, you should also give participants honest answers to any questions they may have.

How do participants respond to being in research and debriefing? A survey of research participants by Janet Brody, John Gluck, and Alfredo Aragon (2000) found that only 32% of research participants surveyed found their research experience completely positive. Participants' reports indicated that the debriefing they received varied in quality, quantity, and format. However, survey respondents reported the most positive debriefing experiences when they were given a thorough explanation of the study in which they had participated and when they were given a detailed account of how the research is broadly relevant. Respondents' biggest complaint about debriefing was that the debriefing was unclear or provided insufficient information.

To summarize, deception raises serious questions about ethical treatment of participants in psychological research. In the absence of alternative techniques, you may find yourself in the position of having to use deception. Strive to maintain the dignity of the participant by using effective debriefing techniques. However, do not be lulled into believing that you can use ethically questionable research techniques just because you include debriefing (Schuler, 1982).

## QUESTIONS TO PONDER

---

1. What is the status of role playing as an alternative to deception?
2. How can you obtain prior consent to be deceived?
3. What is debriefing and how can it be made most effective?
4. What steps can you take to reduce the impact of deception on participants?

## CONSIDERATIONS WHEN USING ANIMALS AS SUBJECTS IN RESEARCH

---

Psychological research is not limited to research with human participants. There is a rich history of using animals as research subjects dating back to the turn of the 20th century. Generally, there is considerable support among psychologists for using animals as subjects in research (Plous, 1996). Plous reports that 80% of respondents to a survey either supported or strongly supported using animals in research. Support for animal research was strongest for research that did not involve suffering, pain, or death of the animals, even if the research was described as having scientific merit and institutional support. Interestingly, respondents were more accepting of animal research involving pain or death for rats or pigeons than for primates or

dogs. Additionally, there is greater support for animal medical research than for animal research directed toward theory testing, cosmetics-safety testing, or agricultural issues (Wuensch & Poteat, 1998). Finally, men tend to be more accepting of animal research than women (Wuensch & Poteat, 1998).

Research using animals must conform to strict federal and local regulations and to ethical guidelines set out by the APA. We discuss these requirements in Chapter 7.

The final section of this chapter considers some factors that become relevant if you decide to use animals as your research subjects.

### Contributions of Research Using Animal Subjects

Animal research has played a prominent role in the development of theories in psychology and in the solution of applied problems. For example, Pavlov discovered the principles of classical conditioning by using animal subjects (dogs). Thorndike laid the groundwork for modern operant conditioning by using cats as subjects. B. F. Skinner developed the principles of modern operant conditioning by using rats and pigeons as subjects.

Snowdon (1983) points out several areas in which research using animal subjects has contributed significantly to knowledge about behavior. For example, animal research has helped explain the variability in behavior across species. This is important because understanding the variability across animal species may help explain the variability in behavior across humans. Also, research using animals has led to the development of animal models of human psychopathology. Such models may help explain the causes of human mental illness and facilitate the development of effective treatments. Animal research also has contributed significantly to explaining how the brain works and how basic psychological processes (such as learning and memory) operate.

### Choosing Which Animal to Use

Animals used in psychological research include (but are not limited to) chimpanzees and gorillas (language-acquisition research), monkeys (attachment-formation research), cats (learning, memory, physiology), dogs (learning, memory), fish (learning), pigeons (learning), and rats and mice (learning, memory, physiology). Of these, the laboratory rat and the pigeon are by far the most popular. The choice of which animal to use depends on several factors. Certain research questions may mandate the use of a particular species of animal. For example, you would probably use chimpanzees or gorillas if you were interested in investigating the nature of language and cognition in nonhuman subjects. In addition, using the same type of animal used in a previous experiment allows you to relate your findings to those previously obtained without having to worry about generalizing across animals.

Your choice of animals also will depend in part on the facilities at your particular institution. Many institutions are not equipped to handle primates or, for that matter, any large animal. You may be limited to using smaller animals such as rats, mice, or birds. Even if you do have the facilities to support the larger animals, your choice may be limited by the availability of certain animals (chimpanzees and monkeys are difficult to obtain). Finally, cost also may be a factor. For example, a cat may cost around \$500 and a monkey over \$1000. Contrast that cost to around \$15 for a laboratory rat.

## QUESTIONS TO PONDER

---

1. What are the general considerations concerning using animals in research?
2. What roles has animal research played in psychology?
3. What factors enter into your decision about which animals to use in your research?

### Why Use Animals?

You might choose to use animals in your research for many reasons. One reason is that some procedures can be used on animals that cannot be used on humans. Research investigating how different parts of the brain influence behavior often uses surgical techniques such as lesions, ablation, and cannula surgery. These procedures obviously cannot be conducted on humans.

As an example, suppose you were interested in studying how lesions to the hypothalamus affect motivation. You probably would not find many humans willing to volunteer for research that involves destroying a part of the brain! Animal subjects are the only available choice for research of this type. Similarly, even if there are areas of research that can be studied with humans (such as examining the effects of stress on learning), you may not be able to expose humans to extremely high levels of an independent variable. Again, animals would be the choice for subjects in research in which the independent variable cannot be manipulated adequately within the guidelines for the ethical treatment of human participants.

In addition to these reasons for choosing animals, animals allow you greater control over environmental conditions (both within the experiment and in the living conditions of the animal). Such control may be necessary to ensure internal validity. By controlling the environment, you can eliminate extraneous, possibly confounding, variables. By using animals, you also have control over the genetic or biological characteristics of your subjects. If you wanted to replicate an experiment that used Long–Evans rats, you could acquire your animals from the same source that supplied them to the author of the original study. Finally, animal subjects are convenient.

### How to Acquire Animals for Research

After you have decided to use animals and have chosen which animals you are going to use, your next step is to acquire the animals. Two methods for acquiring animals are acceptable. First, your institution may maintain a breeding colony. Second, you may use one of the many reliable and reputable breeding farms that specialize in raising animals for research.

Each method has advantages and disadvantages. The on-site colony is convenient, but the usefulness of these animals may be limited. The conditions under which they were bred and housed may cause them to react in idiosyncratic ways to experimental manipulations. Thus, you cannot be sure that the results you produce with on-site animals will be the same as the results that would be obtained had you used animals from a breeding farm.

One advantage to using animals from a breeding farm is that you can be reasonably sure of the history of the animals. These farms specialize in breeding animals

for research purposes. The animals are bred and housed under controlled conditions, ensuring a degree of uniformity across the animals. However, animals of the same strain obtained from different breeding farms may differ significantly. For example, Sprague–Dawley rats obtained from different breeders may differ in subtle characteristics such as reactivity to stimuli. These differences may affect the results of some experiments.

### Generality of Animal Research Data

One criticism of animal research is that the results may not generalize to humans or even to other animal species. This criticism has at its core a basic assumption: All psychological research must be applicable to humans. However, psychology is not concerned only with human behavior. Many research psychologists are interested in exploring the parameters of animal behavior, with little or no eye toward making statements about human behavior.

Much animal research does in fact generalize to humans. The basic laws of classical and operant conditioning, which were discovered through animal research, have been found to apply to human behavior. Figure 6-4 shows a comparison between two extinction curves. Panel (a) shows a typical extinction curve generated by an animal in an operant chamber after reinforcement of a response has been withdrawn. Panel (b) shows the extinction curve generated when a parent stops reinforcing a child's crying at bedtime (Williams, 1959). Notice the similarities. Other examples also can be cited. The effects of alcohol on prenatal development have been studied extensively with rats and mice. The pattern of malformations found in the animal research is highly similar to the pattern observed in the offspring of alcoholic mothers.

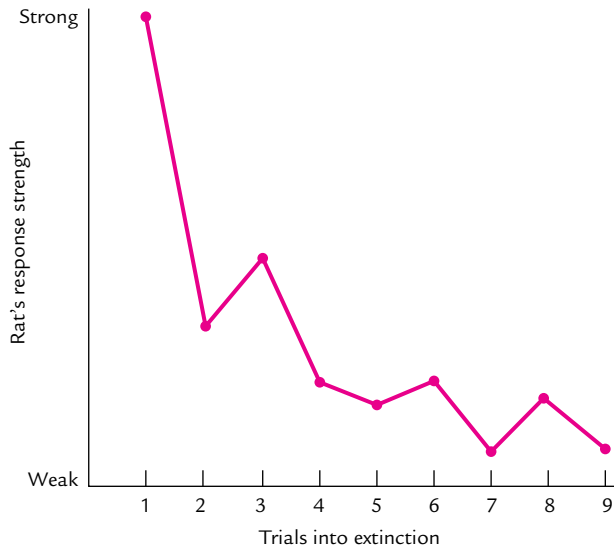
Although results from animal studies often do generalize to humans, such generalization should always be done with caution, as the following example illustrates. In the 1950s, many pregnant women (mainly in Sweden) took the drug thalidomide to help reduce morning sickness. Some of the mothers who took thalidomide gave birth to children with a gross physical defect called *phocomelia*. A child with this defect might be born without legs and have feet attached directly to the lower body. Tests were conducted on rats to determine whether thalidomide was the cause for the malformations. No abnormalities were found among the rats. However, the malformations were found when animals more closely related to humans (monkeys) were used.

Of course, whether results obtained with animal subjects can be applied to humans is an empirical question that can be answered through further research—if the findings have relevance to human behavior, then so much the better. Even if they do not, we gain a better understanding of the factors that differentiate humans from other animals and of the limits to our behavioral laws.

### QUESTIONS TO PONDER

---

1. What arguments can you make for using animals in research?
2. How do you acquire animals for research?
3. What are the main arguments surrounding the generality of animal research data?

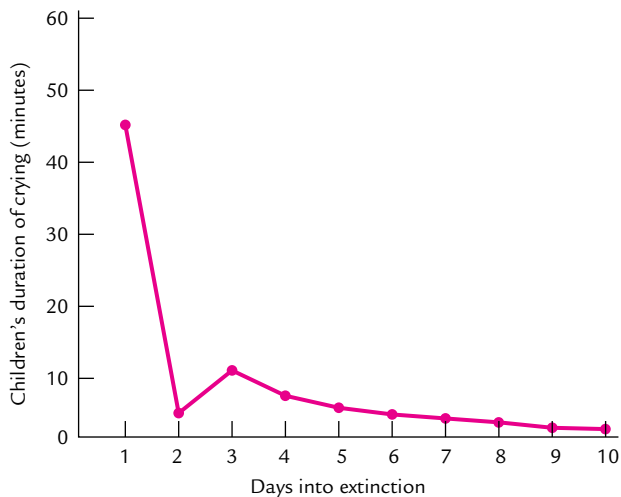


**FIGURE 6-4** Comparison of extinction curves:

(a) a rat's lever-pressing behavior and (b) a child's crying at bedtime.

SOURCE: Panel (b) from Williams, 1959, p. 269; reprinted with permission.

(a)



(b)

## The Animal Rights Movement

Humans have been using animals in research for thousands of years. In fact, we can trace the use of animals in research to coincide with the emergence of medical science (Baumans, 2004). Baumans points out that using animals for medical research goes all the way back to the ancient Greek philosophers such as Aristotle. The Roman physician Galen based many of his medical treatments for humans on physiological experiments for humans on physiological experiments conducted on animals (Baumans, 2004). After a lull in such research in the

Middle Ages, animal experimentation again became popular during the Renaissance period. Philosophers (e.g., Descartes) suggested that animals did not possess a soul or a mind and were basically machines (Baumans, 2004). Trends in the 20th century show that animal research showed a sharp increase between the early 1900s and the 1960s, peaking in 1970 (Baumans, 2004). Baumans reports a small reduction in animal research from its peak to the end of the century. By far, mice and rats make up the majority of animals used in research, accounting for 77% of animals used in research in England (Baumans, 2004) and around 90% in the United States (Shanks, 2003).

Despite the long history of using animals in research, concern has been expressed about using animals in this capacity. Concern over using animals in research stretches back to the early days of using animals (Baumans, 2004). Modern public and political concern over using animals in research can be traced back to the 1874 meeting of the British Medical Association (Matfield, 2002). At the meeting, using a dog as a subject, a doctor demonstrated how an epileptic seizure could be induced with a drug. After the demonstration, some members of the audience protested against using the dog in such a capacity (Matfield, 2002). Organizations to protect animals against cruel treatment, such as the Humane Society and the American Society for the Prevention of Cruelty to Animals (ASPCA) have existed for many years. For example, the ASPCA also dates back to 1874.

The concern over treatment of animals in a variety of areas (farming, research, etc.) has become more visible. People have begun to question seriously the use of animals in research. Many people have taken the position that the role of the animal in research should be reduced. Some have even advocated completely banning the use of animals as subjects in research.

It is important to understand that this issue has potentially serious consequences beyond the moral questions surrounding using animals as research subjects (Shanks, 2003). A majority of research using animals in research is biomedical research (e.g., drug research and testing new medical treatments), which has implications for human health, well-being, and life. A significant reduction in such research may have long-term health consequences for humans. The degree of reduction being advocated varies from a total ban on using animals to simply ensuring that researchers treat their animals ethically.

We can summarize the public and policy debate over using animals in research to two major questions: Is animal research cruel, and is animal research necessary? (Matfield, 2002). Mark Matfield points out that the necessity issue embodies three main points. First, are there viable alternatives to animal research? Second, do results from animal research generalize sufficiently to humans to make it worthwhile? Third, is animal research necessary in general? The remainder of this chapter is devoted to exploring the issues surrounding the arguments made against using animals in research. The intention of this discussion is to present the arguments made by both sides and then analyze them critically. The final judgment about the role of animals in research is left to you.

### Animal Research Issues

Singer (1975, 2002), in a book devoted exclusively to the treatment of animals, raises several objections to using animals in a variety of capacities (from research subjects to food). This discussion is limited to the issue of using animals in research. It is

important to understand Singer's main thesis. Singer (2002) does not maintain that animals and humans are equal in an absolute sense. He does argue, however, that animals and humans are entitled to equal consideration; that differences exist between animals and humans does not justify treating animals in a way that causes suffering. For Singer, the capacity to experience suffering and happiness is central to giving animals equal consideration with humans. Singer (2002) states that it is "speciesist" to give consideration to the pain and suffering of humans but not to animals. According to Singer avoiding speciesism requires an allowance that "all beings who are similar in all relevant aspects have a similar right to life" (Singer, 2002, p. 19). It makes no sense to him that mere membership in the human species grants this right and deprives animals of it.

Within this general philosophical framework, Singer (1975, 2002) maintains that animals should not be used in research that causes them to suffer. Singer (1975) further argues that "most animal studies published are trivial anyway" (p. 227). To support his point, Singer provides a litany of research examples that subjected animals to sometimes painful procedures. Included in this list are the classic studies by Harry Harlow on attachment in infant monkeys and Martin Seligman on learned helplessness. According to Singer, the suffering of the animals was not justified given the trivial nature of the research question and results. Consider an example that Singer (2002) provides (a critical analysis of Singer's assertions follows):

I reported on an experiment performed at Bowling Green University in Ohio by P. Badia and two colleagues, and published in 1973. In that experiment ten rats were tested in sessions that were six hours long, during which frequent shock was "at all times unavoidable and inescapable." The rats could press either of two levers within the test chamber in order to receive a warning of a coming shock. The experimenters concluded that the rats did prefer to be warned of a shock. In 1984 the same experiment was still being carried out. Because someone had suggested that the previous experiment could have been "methodologically unsound," P. Badia, this time with B. Abbott of Indiana University, placed ten rats in electrified chambers, subjecting them again to six-hour shock sessions. . . . The experimenters found, once again, that the rats preferred shock that was signaled, even if it resulted in their receiving more shocks. (Singer 2002, pp. 47–48)

These and several other summaries like them are included in Singer's book to point out the trivial nature of the research results obtained at the expense of animal suffering. If you had read only Singer's book, you would probably come away with the feeling that "everyone already knows that rats will prefer a warning." We can criticize Singer on at least three grounds concerning the brief research summaries.

First, each of the summaries referred to research that was taken out of the theoretical, empirical, or applied context in which the research was originally conducted. By isolating a study from its scientific context, Singer made the research appear trivial. You could take just about any piece of research and trivialize it by removing it from its context. In fact, Badia's studies (summarized in the preceding excerpts) provided important information about how organisms react to stress. To gain a full understanding of the purposes of research, you *must* read the original paper (as pointed out in

Chapter 3). In the introduction to the paper, the author will surely provide the theoretical context and potential importance of the research.

Second, Singer leaves the strong impression that each study of a series merely replicates the ones before it, without contributing anything new to our understanding of the phenomenon under investigation or to its generality across different procedures and contexts. For example, in the paragraph just quoted reviewing the follow-up study by Badia and Abbott (1984), Singer begins by asserting that “the *same* research was still being conducted” (emphasis ours). In fact, scientific understanding of a phenomenon typically progresses by the gradual elimination of rival explanations over a long series of experiments designed for that purpose and by demonstrations that a phenomenon is not an artifact of a particular method of study.

Third, Singer’s presentation of the research strongly suggests that the research was unnecessary because the findings were already obvious and known. Singer committed what social psychologists call the “I-knew-it-all-along phenomenon” (Myers, 1999). The “I-knew-it-all-along phenomenon” refers to the fact that when you hear about some research results, you have the tendency to believe that you already knew that the reported relationship exists. Several researchers (Slovic & Fischhoff, 1977; Wood, 1979) have shown that when individuals are asked to predict the outcome of research *before* they hear the results, they fail. However, when the results are known, they are not surprised. You can demonstrate this for yourself with the following experiment, suggested by Bolt and Myers (1983).

Choose 10 participants for this demonstration. Provide half of them with the following statement:

Social psychologists have found that the adage “Out of sight, out of mind” is valid.

Provide the other half with this statement:

Social psychologists have found that the adage “Absence makes the heart grow fonder” is valid.

Ask participants to indicate whether they are surprised by the finding. You should find that your participants are not surprised by the finding reported to them. Next, have participants write a brief paragraph explaining why they believe that the statement is true. You should find that, in addition to believing that the statement is true, participants will be able to justify the reported finding.

The point of this exercise is that when you are told about the results of research, they often seem obvious. Singer played on this tendency (probably inadvertently) when he presented results from animal studies and then implied that “we knew it all along.” In fact, before the research was done, we probably did *not* know it all along. The research reported by Singer made valuable contributions to science. Taking it out of context and suggesting that the results were obvious leads to the illusion that the research was trivial.

Not all the points made by Singer are invalid. In fact, researchers should treat their animals in a humane fashion. However, you must consider the cost–benefit ratio when evaluating animal research. Is the cost to the subject outweighed by the benefits of the research? Some people within the animal rights movement place a high



value on the cost factor and a low value on the benefit factor. You must consider the benefits of the research that you plan to do on several levels: theoretical, empirical, and applied. In many cases, the benefits derived from the research outweigh the costs to the subjects. You should remember, however, that it is not always immediately obvious what the benefits of a particular line of research might be. It may take several years and a number of studies to be conducted before the benefit of research emerges.

Although controversy over the use of animals in research still exists, the issue may be cooling down. Public opinion of animal research is generally favorable, especially if it is done under the right conditions (Swami, Furnham, & Christopher, 2008). According to a 2005 Hart poll released by the Foundation for Biomedical Research (2005), 76% of Americans polled believed that animal research was important, with 40% indicating it contributed a great deal. Only 14% believed that animal research contributed very little or not at all (Foundation for Biomedical Research, 2005). The poll also showed that 56% believed that current regulations are sufficient to protect animals used in research. There also appear to be some differences across nationalities and between genders. For example, Swami, Furnham, and Christopher (2008) found that Americans held more positive attitudes toward animal testing and less concern for animal welfare than individuals from Great Britain. They also found that women were more strongly against animal testing than were men.

The tensions between animal rights activists and researchers may also be lessening. A study by Plous (1998) compared attitude changes of animal rights activists between 1990 and 1996. Plous reports that in 1990 a majority of animal rights activists believed that using animals in research was the most important issue facing the animal rights movement. A similar survey of activists done in 1996 revealed that a majority of activists believed that the use of animals in agriculture was the number-one issue facing the animal rights movement. Further, respondents to the 1996 survey advocated less radical methods for dealing with animals used in research. For example, fewer respondents (compared to the 1990 survey) advocated break-ins at laboratories using animals as a method of controlling the use of animals in research. In fact, most respondents in 1996 advocated more dialogue between activists and animal researchers.

We think it is important that you understand that those who advocate for animal rights are not bad people. Quite the contrary, typically individuals who advocate animal rights have a genuine interest in protecting the welfare of animals. In fact, such individuals have a high level of moral reasoning (Block, 2003), have positive attitudes concerning animal welfare (Signal and Taylor, 2006), hold romantic views of the environment (Kruse, 1999), and even have dreams with more animal characters than the general population (Lewis, 2008).

Having said this, we should note that some extreme animal rights activists resort to radical tactics. For example, on November 14, 2004, members of the Animal Liberation Front (ALF) broke into the psychology department's animal laboratory at the University of Iowa. According to their press release, they "liberated" 88 mice and 313 rats. According to university officials, the ALF activists also destroyed up to 30 computers and poured acid on equipment. University of Iowa President David Skorton testified before Congress that the break-in caused about \$450,000 in damage (GazetteOnline, 2005). It may be a while before common ground can be found.

## ALTERNATIVES TO ANIMALS IN RESEARCH: IN VITRO METHODS AND COMPUTER SIMULATION

---

Animal rights activists point out that viable alternatives to using living animals in research (known as *in vivo* methods) exist, two of which are *in vitro* methods and computer simulations. These methods are more applicable to biological and medical research than to behavioral research. *In vitro* (which means “in glass”) *methods* substitute isolated living tissue cultures for whole, living animals. Experiments using this method have been performed to test the toxicity and mutagenicity of various chemicals and drugs on living tissue. *Computer simulations* also have been suggested as an alternative to using living organisms in research. In a computer simulation study, a mathematical model of the process to be simulated is programmed into the computer. Parameters and data concerning variables fed into a computer then indicate what patterns of behavior would develop according to the model.

Several problems with *in vitro* and computer simulation methods preclude them from being substitutes for psychological research on living organisms. In drug studies, for example, *in vitro* methods may be adequate in the early stages of testing. However, the only way to determine the drug’s effects on behavior is to test the drug on living, behaving animals. At present, the behavioral or psychological effects of these chemical agents cannot be predicted by the reactions of tissue samples or the results of computer simulations. Behavioral systems are simply too complex for that. Would you feel confident taking a new tranquilizer that had only been tested on tissues in a petri dish?

The effects of environmental variables and manipulations of the brain also cannot be studied using *in vitro* methods. It is necessary to have a living organism. For example, if you were interested in determining how a particular part of the brain affects aggression, you could not study this problem with an *in vitro* method. You would need an intact organism (such as a rat) in order to systematically manipulate the brain and observe behavioral changes.

A different problem arises with computer simulation. You need enough information to write the simulation, and this information can only be obtained by observing and testing live, intact animals. Even when a model has been developed, behavioral research on animals is necessary to determine whether the model correctly predicts behavior. Far from eliminating the need for animals in behavioral research, developing and testing computer simulations actually increases this need.

In short, there are really no viable alternatives to using animals in behavioral research. Ultimately, it is up to you to be sure that the techniques you use do not cause the animals undue suffering. Always be aware of your responsibility to treat your animal subjects ethically and humanely.

## QUESTIONS TO PONDER

---

1. What basic arguments do animal rights activists make concerning the use of animals in research?
2. What are Singer’s criticisms of animal research?
3. What arguments can be made against Singer’s views of animal research?

4. What evidence can you cite that the animal rights controversy might be settling down, or perhaps not settling down?
5. What alternatives have been proposed to using animals in research and why do some of them not apply to behavioral sciences?

## SUMMARY

---

After you have developed your research idea into a testable hypothesis and settled on a research design, your next step is to recruit participants or subjects for your study. Before you can proceed with your study, however, it must be evaluated for ethical issues. A review board will determine if your research protocol adheres to accepted ethical guidelines. Before you begin your research there are several issues you must consider when using human participants or animal subjects in your research. One important general consideration is the sample you will use in your research. It is not practical to include all members of a given population (e.g., third-grade children, college students) in your research. Instead you select a smaller sample of the population to include in your research.

One goal is to generalize the findings from your sample used in your study to the larger population. This is most effectively accomplished when you have a random sample of participants in your study, meaning that each individual in the population has an equal chance of being selected for inclusion in your sample. The reality of psychological research is that the ideal of a random sample is rarely achieved. Instead, nonrandom samples are used because they are convenient. In many psychological studies college students are used because they comprise the subject pools at many universities. Nonrandom samples are also common in studies conducted on the Internet and in animal research. Using subjects obtained through nonrandom sampling may limit the generality of your results. However, there are situations in which random sampling may not be necessary.

Regardless of the type of research you conduct using human participants, you must consider three factors: the setting in which your research will take place (field or laboratory), any special needs of your particular research (e.g., needing participants with certain personality characteristics), and any institutional, departmental, and ethical policies and guidelines governing the use of participants in research.

The requirement of voluntary participation and full disclosure of the methods of your research may lead to problems. For example, individuals who volunteer have been found to differ from nonvolunteers in several ways. This volunteer bias represents a threat to both internal and external validity. It can be counteracted to some extent by careful participant recruitment procedures. In cases in which you must use a deceptive technique, take special care to ensure that your participants leave your experiment in the proper frame of mind. You can accomplish this through using role playing or using effective debriefing techniques. At all times, however, you must remain cognizant of the problems with deception even if debriefing is used.

A large amount of psychological research uses animal subjects. Animals are preferred to humans in situations in which experimental manipulations are unethical for use with humans. In recent decades, the animal rights movement has evolved to

challenge the use of animals in research. Animal rights advocates push for restricting the use of animals in research and call for ethical treatment. However, if you use animal subjects, you are still bound by a strict ethical code. Animals must be treated humanely. It is to your advantage to treat your animals ethically because research shows that mistreated animals may yield data that are invalid.

Alternatives to using animals in research have been proposed, including the use of in vitro testing and computer simulation. These alternatives, unfortunately, are not viable for behavioral research in which the goal is to understand the influences of variables on the behavior of the intact, living animals.

### KEY TERMS

---

population

sample

generalization

random sample

nonrandom sample

volunteer bias

deception

role playing

debriefing

# Understanding Ethical Issues in the Research Process

As characterized in Chapter 1, the research process involves a regularized progression of getting and developing research ideas, choosing a research design, deciding on a subject population to use, conducting your study, analyzing data, and reporting results. Central to research in the social sciences in general and psychology in particular is the inclusion of living organisms as research subjects.

Using living organisms, whether human or animal, in research imposes upon you an obligation to treat those organisms in a humane, respectful, and ethical manner. In this chapter, we review various aspects of ethics as they apply to the research process. We explore the ethical issues that apply to research using human participants, including a brief history of the evolution of the ethical principles that guide research with human participants. We also explore the ethical principles that apply to using animal subjects in research. Finally, we explore another issue of research ethics: your obligation as a researcher to approach your science ethically and honestly.

## ETHICAL RESEARCH PRACTICE WITH HUMAN PARTICIPANTS

In the early years of psychological research, researchers were pretty much left on their own to conduct their research. They decided when, how, and with whom research would be conducted. Little, if any, attention was paid to ethical issues. Researchers were responsible for making their own determinations about ethical research practice. Unfortunately, this led to some experiments that would most likely be considered unethical by today's standards. Let's look at a couple of examples.

### John Watson and Little Albert

John Watson was the founder of the behaviorist school of psychology. According to behaviorism, the subject matter of psychology was observable stimuli (S) and observable responses (R). One of

## CHAPTER OUTLINE

### Ethical Research Practice With Human Participants

John Watson and Little Albert

Is It Fear or Is It Anger?

Putting Ethical Considerations in Context

The Evolution of Ethical Principles for Research With Human Participants

Nazi War Crimes and the Nuremberg Code

The Declaration of Helsinki

The Belmont Report

APA Ethical Guidelines

Government Regulations

Internet Research and Ethical Research Practice

Ethical Guidelines, Your Research, and the Institutional Review Board

### Ethical Considerations When Using Animal Subjects

The Institutional Animal Care and Use Committee

Cost-Benefit Assessment: Should the Research Be Done?

### Treating Science Ethically: The Importance of Research Integrity and the Problem of Research Fraud

What Constitutes Fraud in Research?

The Prevalence of Research Fraud

Explanations for Research Fraud

Dealing With Research Fraud

### Summary

### Key Terms

Watson's studies attempted to determine if emotional responses could be learned. He along with a graduate student named Rosalie Rayner conducted a study in which a young child (Albert) was exposed to a white rat. Initially, Albert showed no negative response to the white rat. Next, Watson and Rayner (1920) presented Albert with the white rat followed by a loud noise produced by striking a steel bar with a hammer behind Albert. After several instances in which the white rat and clanging of the steel bar were presented jointly, Watson and Rayner tested Albert's reaction to the white rat alone. Here is what they found:

Rat alone. The instant the rat was shown the baby began to cry. Almost instantly he turned sharply to the left, fell over on left side, raised himself on all fours and began to crawl away so rapidly that he was caught with difficulty before reaching the edge of the table. (Watson & Rayner, 1920, p. 5)

Watson and Rayner (1920) continued their study by testing Albert's reactions to a number of other stimuli (a white rabbit, some toy blocks, and a fur coat) and found that Albert showed a negative reaction to stimuli that were similar to the white rat (the rabbit or fur coat), but not toward other stimuli (the blocks). They concluded that Albert's negative conditioned emotional response had transferred to the other similar stimuli. Finally, Watson and Rayner wanted to study "detachment" of the conditioned emotional response to the white rat. That is, they wanted to see if they could eliminate or reduce the negative emotional response that they had conditioned into Albert. Unfortunately, Albert's mother (who worked at the hospital where the experiment was being conducted) left the hospital, taking Albert with her. Watson and Rayner never got to reverse the conditioning process in their lab.

**Ethical Issues Raised by the Watson and Rayner Study** Do you see any ethical issues or problems raised by Watson and Rayner's study? Do you think that you could conduct this same study today? Let's review this study and identify some ethical issues it raises. First, Watson and Rayner make no mention of whether Albert's mother granted permission to use Albert in their study. This certainly raises the important issue of consent. Current research practice, as we explain later in this chapter, requires obtaining **informed consent**, a process that involves informing a participant about research and obtaining consent to participate in it. The participant reads and signs a form specifying the purpose of a study, the methods to be used, the requirements for participation, the costs and benefits of research participation, that participation is voluntary, and that the participant is free to withdraw from the research at any time without penalty. It is especially important to obtain informed consent when the participant is a minor child. In Albert's case his mother should have been provided with informed consent. Second, one can legitimately question whether it is ethical to condition fear into an 11-month-old child. What short-term and/or long-term consequences could the conditioning process have had on Albert's behavior and well-being? Third, Watson and Rayner were unable to reverse the effects of the conditioning process because Albert's mother removed him from the hospital. It would be incumbent upon any modern researcher to remove any ill effects of the experimental manipulations.

## Is It Fear or Is It Anger?

For many years, psychologists have wondered about the physiological underpinnings of emotions. Does each emotion have its own, unique physiological response? Or do all emotions share a common physiological response? An experiment conducted by Albert Ax (1953) sought to address these questions. Ax obtained physiological data from participants who were induced to experience the emotions of fear and anger. It is how Ax induced the fear that might raise some ethics eyebrows.

Participants were told that they were taking part in an experiment to study the physiological differences between people with and without hypertension. Ax hooked participants up to a shock generator and gave them a series of mild electric shocks that did not cause any pain. Ax then instructed the participant to indicate when the shock first could be felt. When the participant reported feeling the shock, the experimenter expressed surprise and proceeded to check the wiring on the shock generator. While checking the wiring, the experimenter secretly pressed a switch that caused the shock generator to spew sparks near the participant. At this point, the experimenter, in an alarmed voice, said that there was a “dangerous high voltage short circuit” (Ax, 1953, p. 435). After 5 minutes, the experimenter removed the wire from the participant’s finger, telling the participant that there was no longer any danger.

**Ethical Issues Raised by Ax’s Study** How would you have felt if you had been in Ax’s experiment and been subjected to the fear-inducing procedure? Would you have felt that your life was in danger or that you could be seriously harmed? Of course, in Ax’s procedure, participants were not actually in any danger. The sparking and reactions from the technician were all staged, but the participant did not know that. The biggest ethical question surrounding Ax’s procedure is the use of deception. Whenever you tell your participants something that is false (or withhold information), you are using deception. Is it ethical to lie to people in the name of science? Should there have been full disclosure of the procedure for inducing fear before the experiment? These are questions that must be addressed when you consider using a deceptive research procedure. In fact, deception is addressed by the ethical guidelines of the American Psychological Association (APA), which we discuss later in the chapter.

## Putting Ethical Considerations in Context

The Watson and Rayner and the Ax studies illustrate some of the ethical issues that arise when you do research. Currently there are numerous rules, regulations, and guidelines regarding research ethics that you must follow. You must present your research protocol for review of ethical issues *before* you can conduct your research. Your proposal is reviewed to make sure that the safety, well-being, dignity, and rights of your participants are protected.

The rules that define ethical research practice did not emerge overnight. Instead, they evolved slowly over a number of years and in reaction to various ethical issues that emerged along the way. In the next section, we review the evolution of the present-day ethical guidelines that apply to research using human participants. In a later section, we explore the ethical guidelines that apply to research with animal subjects.

## QUESTIONS TO PONDER

---

1. What ethical issues does Watson and Rayner's "Little Albert" study raise?
2. What ethical issues does Ax's experiment raise?
3. What could you do to address some of the ethical issues raised in the two studies reviewed?

### The Evolution of Ethical Principles for Research With Human Participants

In 1954 W. Edgar Vinacke wrote a letter to the editor of the *American Psychologist* (the official journal of the APA) taking psychologists to task for a lack of concern over the welfare of their research participants. In his letter, Vinacke pointed out that the psychological researcher frequently misinforms participants (as in Ax's study) or exposes them to painful or embarrassing conditions (as in Watson and Rayner's study), often without revealing the nature and purpose of the study.

Although Vinacke's concerns were well founded and represented some of the earliest criticisms of research practice among psychologists, the concern over ethical treatment of research participants predates Vinacke's letter by several years. The APA established a committee in 1938 to consider the issue of ethics within psychological research (Schuler, 1982).

The current concern over ethical treatment of research participants can be traced to the post-World War II Nuremberg war crimes trials. Many of the ethical principles eventually adopted by the APA in 1951 are rooted in what is now called the Nuremberg Code.

### Nazi War Crimes and the Nuremberg Code

In the years before World War II the Nazis enacted several anti-Jewish laws (laws preventing Jews from holding civil service jobs, shopping in non-Jewish stores, etc.) and promoted virulent prejudice against Jews. Through shrewd propaganda, the Nazis were able to convince the public (albeit incorrectly) that Jews were the cause of the "ills" that befell the German people after World War I. As a result of these laws, a number of concentration camps and death camps were established to which millions of Jews were deported. Many of these concentration camps served as slave-labor camps. Others (Auschwitz, Treblinka, and Sobibor) had another purpose: to carry out Hitler's "final solution of the Jewish problem." The principal reason for the existence of this latter group of camps was the systematic extermination of human beings.

At Auschwitz "medical" experiments were conducted on some of the doomed inmates. For example, an SS doctor at Auschwitz named Josef Mengele selected inmates at "the ramp" for either immediate extermination or incarceration in the camp as the inmates arrived at the camp. Some of those spared (most notably twins) served as participants in a variety of experiments. Some of the experiments were carried out in the name of eugenics and were aimed at proving the existence of a master race or "improving" the genetic stock of such a race. Mass sterilization procedures (without anesthesia) were tried out on inmates in an attempt to find



the most efficient way to reduce the population of “inferior races.” Other experiments were carried out for the German military. For example, inmates were placed in decompression chambers to see how long it would take them to die under high-altitude conditions or were immersed in near-freezing water to see how long a pilot could survive in the water before rescue (research carried out for the German Air Force). Bones were broken and rebroken to see how many times they could be broken before healing was not possible. The list of these sadistic “experiments” goes on and on.

In all of these experiments the inmates were unwilling participants. They certainly did not freely volunteer and give their free consent to be participants in this cruel research. After the war, when the Nazi atrocities became known, some of those responsible were tried for their crimes at the Nuremberg trials. A special “Doctor’s Trial” put on trial Nazi physicians who participated in the heinous medical experiments (unfortunately, Mengele escaped and was not tried). It became evident as the trial progressed that there was no clear statement about ethical medical research practice (Cohen, Bankert, & Cooper, 2005). A majority of the doctors who were tried were convicted. As important, however, out of the trials came the **Nuremberg Code**, which laid the groundwork for many of the current ethical standards for psychological and medical research. The 10 major principles set forth in the Nuremberg Code (Katz, 1972, pp. 305–306) are listed in Table 7-1.

Note that Point 1 requires that participation in research be voluntary and that the participant has the right to know about the nature, purposes, and duration of the research. In addition, Points 2 and 3 suggest that frivolous research is unethical. Scientists should not subject people to experimental manipulations if there is another way to acquire the same information, and a firm scientific base must exist for the experiment. Points 4 to 8 place the responsibility on the researcher to ensure that participants are not exposed to potentially harmful research practices. Finally, Points 9 and 10 require that research be terminated by either the participant or experimenter if it becomes obvious to either that continuation of the experiment would be, for any reason, unacceptable. These factors were embodied in the ethical standards adopted by the APA and the U.S. Department of Health and Human Services (HHS).

### The Declaration of Helsinki

Another major step in the evolution of ethical codes came in 1964 when the **Declaration of Helsinki** was adopted by the World Medical Association. Although the Declaration of Helsinki specifically addressed medical research, it embodied many principles that also apply to research in the social sciences. For example, one of the basic principles is that medical researchers are obligated to protect the health, welfare, and dignity of research participants. Another basic principle states that all medical research must conform to accepted scientific principles and be based on knowledge of relevant scientific literature. The declaration also states that research must be reviewed by an independent group of individuals who will ensure that the research protocol adheres to accepted ethical standards. As we will see below, all of these principles are embodied in the code of ethics adopted by the APA.

**TABLE 7-1 Ten Points of the Nuremberg Code**

1. Participation of subjects must be totally voluntary and the subject should have the capacity to give consent to participate. Further, the subject should be fully informed of the purposes, nature, and duration of the experiment.
2. The research should yield results that are useful to society and that cannot be obtained in any other way.
3. The research should have a sound footing in animal research and be based on the natural history of the problem under study.
4. Steps should be taken in the research to avoid unnecessary physical or psychological harm to subjects.
5. Research should not be conducted if there is reason to believe that death or disability will occur to the subjects.
6. The risk involved in the research should be proportioned to the benefits to be obtained from the results.
7. Proper plans should be made and facilities provided to protect the subject against harm.
8. Research should be conducted by highly qualified scientists only.
9. The subject should have the freedom to withdraw from the experiment at any time if he or she has reached the conclusion that continuing in the experiment is not possible.
10. The researcher must be prepared to discontinue the experiment if it becomes evident to the researcher that continuing the research will be harmful to the subjects.

SOURCE: Based on Katz, 1972, pp. 305-306.

## The Belmont Report

The **Belmont Report** was issued in 1979 and further delineated ethical research practice with human participants (Cohen et al., 2005). The Belmont Report was issued to clarify the information used by members of the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research who adopted the National Research Act in 1974. The Belmont Report presents three basic principles of ethical treatment of human participants underlying all medical and behavioral research: respect for persons, beneficence, and justice (Belmont Report, 1979). Several of the principles elaborated below have been incorporated into ethical codes developed by professional organizations, including the American Psychological Association.

1. **Respect for persons.** Respect for persons involves two components. First, research participants must be treated as autonomous persons who are capable of making their own decisions. Second, persons with diminished autonomy or capacity deserve protection. On a practical level, this provision requires that research participants enter into participation voluntarily and be fully informed.

2. **Beneficence.** Ethical research practice not only requires respect for persons but also includes a requirement to protect the well-being of research participants. Beneficence includes two components: to do no harm to participants and to maximize benefits while minimizing harm.
3. **Justice.** The principle of justice divides the burden of research equally between the researcher and the participant. Each should share in the costs and potential benefits of the research. The principle of justice also proscribes using participant populations simply because they are readily available, are convenient, and may have difficulty refusing participation in research.

## QUESTIONS TO PONDER

---

1. What is the Nuremberg Code, and how does it relate to current ethical guidelines?
2. What did the Declaration of Helsinki add to the Nuremberg Code?
3. What are the three principles laid out in the Belmont Report?

## APA Ethical Guidelines

The APA began preparing its ethical guidelines in 1947. Complaints from members of the APA served as the impetus for looking into the establishment of ethical guidelines for researchers. The first ethical code of the APA was accepted in 1953 (Schuler, 1982). Since their original publication in 1953, the APA guidelines have been revised several times, most recently in 2002, which took effect in June 2003. The APA's **Ethical Principles of Psychologists and Code of Conduct 2002** is a comprehensive document specifying the ethical responsibilities of psychologists and researchers. The document is too long to present in its entirety. Table 7-2 presents the most recent version of the guidelines for using human participants in research. Review the points of the Nuremberg Code shown in Table 7-1 and the three principles of the Belmont Report and note elements of those documents are reflected in the current APA guidelines. The APA (1973) also has established a set of ethical guidelines for research in which children are used as participants. If you are going to use children as participants, you should familiarize yourself with those guidelines.

## Government Regulations

The period spanning the early 1940s through the late 1950s was one in which researchers became increasingly concerned with the ethical treatment of research participants. This was true for researchers in psychology as well as in the medical profession. However, despite the Nuremberg Code, Helsinki Declaration, Belmont Report, and APA ethical guidelines research, abuses continued. The greater sensitivity about ethics and the newly drafted guidelines did not ensure that research was carried out in an ethical manner, as the next example clearly shows.

The director of medicine at the Jewish Chronic Disease Hospital in Brooklyn, New York, approved the injection of live cancer cells into two chronically ill patients

**TABLE 7-2 Summary of the 2002 APA Ethical Principles That Apply to Human Research Participants**

1. Research proposals submitted to Institutional Review Boards shall contain accurate information. Upon approval researchers shall conduct their research within the approved protocol.
2. When informed consent is required, informed consent shall include: (1) the purpose of the research, expected duration, and procedures; (2) their right to decline to participate and to withdraw from the research once participation has begun; (3) the foreseeable consequences of declining or withdrawing; (4) reasonably foreseeable factors that may be expected to influence their willingness to participate such as potential risks, discomfort, or adverse effects; (5) any prospective research benefits; (6) limits of confidentiality; (7) incentives for participation; and (8) whom to contact for questions about the research and research participants' rights. They provide opportunity for the prospective participants to ask questions and receive answers.
3. When intervention research is conducted that includes experimental treatments, participants shall be informed at the outset of the research of (1) the experimental nature of the treatment; (2) the services that will or will not be available to the control group(s) if appropriate; (3) the means by which assignment to treatment and control groups will be made; (4) available treatment alternatives if an individual does not wish to participate in the research or wishes to withdraw once a study has begun; and (5) compensation for or monetary costs of participating including, if appropriate, whether reimbursement from the participant or a third-party payor will be sought.
4. Informed consent shall be obtained when voices or images are recorded as data unless (1) the research consists solely of naturalistic observations in public places, and it is not anticipated that the recording will be used in a manner that could cause personal identification or harm, or (2) the research design includes deception, and consent for the use of the recording is obtained during debriefing.
5. When psychologists conduct research with clients/patients, students, or subordinates as participants, psychologists take steps to protect the prospective participants from adverse consequences of declining or withdrawing from participation. When research participation is a course requirement or an opportunity for extra credit, the prospective participant is given the choice of equitable alternative activities.
6. Informed consent may be dispensed with only (1) where research would not reasonably be assumed to create distress or harm and involves (a) the study of normal educational practices, curricula, or classroom management methods conducted in educational settings; (b) only anonymous questionnaires, naturalistic observations, or archival research for which disclosure of responses would not place participants at risk of criminal or civil liability or damage their financial standing, employability, or reputation, and confidentiality is protected; or (c) the study of factors related to job or organization effectiveness conducted in organizational settings for which there is no risk to participants' employability, and confidentiality is protected or (2) where otherwise permitted by law or federal or institutional regulations.

**TABLE 7-2 Summary of the 2002 APA Ethical Principles That Apply to Human Research Participants** *continued*

7. Psychologists make reasonable efforts to avoid offering excessive or inappropriate financial or other inducements for research participation when such inducements are likely to coerce participation. When offering professional services as an inducement for research participation, psychologists clarify the nature of the services, as well as the risks, obligations, and limitations.
8. Deception in research shall be used only if they have determined that the use of deceptive techniques is justified by the study's significant prospective scientific, educational, or applied value and that effective nondeceptive alternative procedures are not feasible. Deception is not used if the research is reasonably expected to cause physical pain or several emotional distress. Psychologists explain any deception that is an integral feature of the design and conduct of an experiment to participants as early as is feasible, preferably at the conclusion of their participation, but no later than at the conclusion of the data collection, and permit participants to withdraw their data.
9. (a) Psychologists offer participants a prompt opportunity to obtain appropriate information about the nature, results, and conclusions of the research, and they take reasonable steps to correct any misconceptions that participants may have of which the psychologists are aware. (b) If scientific or humane values justify delaying or withholding this information, psychologists take reasonable measures to reduce the risk of harm. (c) When psychologists become aware that research procedures have harmed a participant, they take reasonable steps to minimize the harm.

SOURCE: APA, 2002.

in July 1963. The patients were unaware of the procedure, which was designed to test the ability of the patients' bodies to reject foreign cells (Katz, 1972). Predictably, the discovery of this ethical violation of the patients' rights raised quite a controversy.

Because of abuses similar to the Jewish Chronic Disease Hospital case, the U.S. government addressed the issue of ethical treatment of human participants in research. The result of this involvement was the establishment of the HHS guidelines for the "protection of human subjects" (U.S. Department of Health and Human Services, 2009). These guidelines specify which categories of research must be reviewed by an institutional review board, and the rules under which review board approval shall be granted. You can find the HHS guidelines at <http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.htm>.

There are also guidelines that apply to using children as research participants. The U.S. Department of Health and Human Services (2009) regulations for research with human participants states that unless the research involving children is exempt under the code, the assent of the child must be obtained. This means that the child must be informed about the study and must give his or her permission for participation. If the child is not old enough to give such assent, then permission must be

obtained from one or both parents. Permission from one parent is sufficient if the research poses no more than minimal risk or has a direct potential benefit to the child participant. Permission from both parents is required if there is greater than minimal risk and there is no direct benefit to the child participant.

The federal regulations covering the use of human research participants and the APA code of ethics are both intended to safeguard the health and welfare of child research participants. However, ethical issues arise even in cases in which all regulations and codes are followed. Take the case of memory-implantation research conducted with children. In a typical experiment, an event that never happened will be implanted in a child's memory. The purpose of this type of research is to discover the extent to which memories can be implanted in children. Douglas Herrmann and Carol Yoder (1998) have raised some serious ethical issues concerning this type of research. They point out that children and adults may respond very differently to the deception involved in implanted-memory research. They argue further that children may not fully understand the nature of the deception being used and may be participating only under parental permission. Herrmann and Yoder suggest that at the time parental permission is sought, it is not possible to fully inform parents of the potential risks because those risks are not fully understood. They called upon researchers in this area to rethink the ethics of the implanted-memory procedure with children.

On the other side of the argument, Stephen Ceci, Maggie Bruck, and Elizabeth Loftus (1998), while agreeing that it is important to protect the welfare of child participants, state that many of the risks that Herrmann and Yoder wrote about were either inflated or nonexistent. In addition, one must also balance the potential risk to the individual child against the potential benefits that come from systematic research (Ornstein & Gordon, 1998). However, Ornstein and Gordon point out that it is essential for researchers to follow up with parents to make sure that child participants do not experience negative side effects because of their participation in a memory-implantation study. They also suggest that careful screening of children (for psychopathology and self-esteem) be conducted before allowing a child to participate.

As you can see, issues surrounding using children as research participants are complex. There is no simple answer to the question of whether children should be allowed to participate in psychological experiments. Certainly, it is important to protect the welfare of children who take part in research. However, discontinuing an important line of research with potential benefits to society would be “throwing the baby out with the bathwater.”

## QUESTIONS TO PONDER

---

1. What are the main points of the APA code of research ethics?
2. What guidelines were instituted by the Department of Health and Human Services, and why were they necessary?
3. What are the ethical issues raised by using children as research participants?

## Internet Research and Ethical Research Practice

The Internet provides researchers with a new way to conduct research. Using the Internet to conduct research raises questions concerning how ethical guidelines developed for offline research apply to research conducted on the Internet. In some cases, ethical guidelines transfer quite well. Some Internet research involves a potential participant going to a Web site and choosing a study to participate in. As in an offline study, the participants will be given a full description of the study, an informed-consent form, and an opportunity to withdraw from the study. They will also receive information on how to obtain follow-up information. This category of research poses no more ethical concerns than offline research.

Another form of Internet research involves issues not covered well by existing ethical guidelines. Research using existing chat rooms, online communities, e-mail groups, or listserves falls into this category. For example, entering a chat room to study the interactions among the participants raises two ethical issues. First, how do you obtain informed consent from the chat room participants? Second, how do you protect the privacy and confidentiality of research participants online? Should participants who agree to remain in the chat room be assigned pseudonyms to protect their identities?

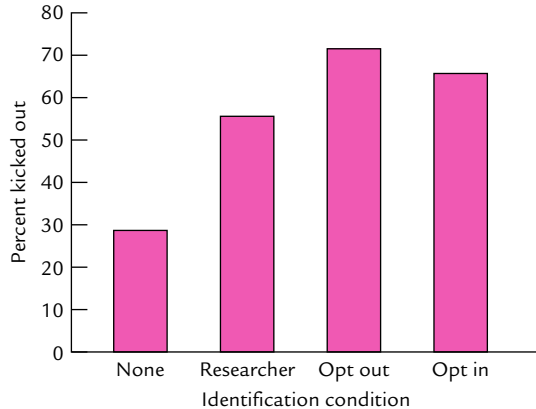
**Informed-Consent Issues** Resolving the issue of informed consent would seem a simple matter: Just have willing participants sign an electronic version of a consent form. However, this procedure, which works well in other contexts, may not work well when studying chat room dynamics (Jones, 1994). Robert Jones questions whether it is ethical to exclude individuals from a chat room (especially if it is one that people join to get help for some condition) if they refuse to take part in the research study. One solution would be to allow everyone to participate but exclude responses from those who refuse to be in the study. Jones, however, questions whether this is feasible and whether it is possible to ensure the anonymity and privacy of nonparticipants.

How might chat room participants respond to being part of a research study? James Hudson and Amy Bruckman (2004) investigated this question, and the results were not pretty. Hudson and Bruckman entered more than 500 chat rooms under one of four conditions. In the first condition, they did not identify themselves as researchers conducting a study. In the second condition, they entered the chat room and identified themselves as researchers studying language use in chat room discussions. In the third condition, they entered the chat room, identified themselves as researchers, and gave chat room participants the option of privately opting out of the study. The fourth condition was identical to the third except that chat room participants were given an opportunity to privately opt into the study.

As shown in Figure 7-1, the researchers were more likely to be kicked out of a chat room when they identified themselves as researchers. It didn't matter which introduction they used. Additionally, they found that they were more likely to be kicked out of small chat rooms than large ones. Hudson and Bruckman (2004) note that for every increase of 13 chat room members, the likelihood of being kicked out was halved. As the number of moderators in a chat room increased, so did the likelihood of being kicked out. Hudson and Bruckman's results indicate that in general chat

**FIGURE 7-1** Influence of researcher identification method on the percentage of times that the researcher is kicked out of an online chat room.

SOURCE: Based on data from Hudson and Bruckman, 2004.



room members do not like being studied. This can pose serious problems for Internet researchers who must obtain informed consent from potential research participants.

**Privacy and Confidentiality Issues** With respect to the privacy and confidentiality issue, David Pittenger (2003) points out that the Internet, by its very nature, is a very public medium. The public nature of the Internet poses a serious threat to the privacy of Internet research participants (Pittenger, 2003). Pittenger raises two concerns.

The first concern is a technical one and refers to the protections that are available in the software programs used by researchers. These programs vary in their ability to secure unauthorized access by computer hackers. You must be reasonably sure that hackers will not gain access to participants' responses. Additionally, if data are housed on a publicly owned computer, the data may be vulnerable to exposure by existing freedom of information laws (Pittenger, 2003). You are ethically bound to protect the confidentiality of participants' responses. You can do this by using appropriate hardware and software and by keeping data stored on a portable storage device like a CD-ROM or memory stick. Of course, you should inform potential participants of the possibility of data disclosure.

The second concern is over the ethical responsibilities of researchers who insinuate themselves into online groups (e.g., chat rooms and communities) without identifying themselves as researchers. You, as the researcher, must be mindful of whether the group that you are studying is a public or private group (Pittenger, 2003). Entering a private group poses special ethical concerns for your research. Research on participants in public groups may pose fewer concerns. Pittenger offers three arguments for considering the Internet equivalent to a public place like a shopping mall.

1. Internet use is now so common that users should understand that it does not afford privacy.
2. A person can easily maintain anonymity by using a pseudonym that cannot be traced back to reveal the user's identity.
3. The exchange of information in open, public Internet forums does not fall under the heading of research that requires informed consent and can be legitimately studied as long as there is no potential harm to participants.



Of course, such arguments would not apply to forums or groups that are advertised as being “confidential” or as having limited access. Internet groups that are created for people with medical conditions (e.g., AIDS) or other afflictions (e.g., alcoholism) often include these provisions. Doing research on such groups would require a more strict research protocol including full disclosure and informed consent.

Pittenger (2003) suggests the following guidelines for ethical research on Internet groups:

1. Learn about and respect the rules of the Internet group that you are going to study. Find out if the group is an open, public one or if it is a closed, private one.
2. Edit any data collected. Comb through the data that you collect and eliminate any names or pseudonyms that may lead to participant identification. You should also eliminate any references to the name of the group being studied.
3. Use multiple groups. You might consider studying several groups devoted to the same topic or issue (e.g., several groups for alcoholics). In addition to increasing the generality of your results, this technique adds another layer of protection to participant privacy.

**Deception in Internet Research** The APA ethical guidelines permit deceptive research under certain conditions. In and of itself, deception does not automatically qualify research as unethical. However, you must be especially careful to protect the dignity of research participants if you use deception. When deception is used, you have an obligation to debrief your participants and dehoax them (see Chapter 6). *Debriefing* means that you explain the methods used in your study, including any deception. *Dehoaxing* means that you convince participants that the deception was necessary and take steps to reverse any ill effects of being deceived. Pittenger (2003) suggests that debriefing and dehoaxing may be more difficult in Internet research. If, for example, participants leave a group before the end of a session or the entire study, it may be difficult to track them down for debriefing and dehoaxing. Pittenger suggests creating a separate Internet group or “enclave” where participants can go for debriefing and dehoaxing.

## QUESTIONS TO PONDER

---

1. What special ethical concerns face you if you conduct your research on the Internet?
2. What are the issues involved in obtaining informed consent in Internet research?
3. What are the issues surrounding privacy and confidentiality in Internet research?
4. What steps can be taken to protect Internet participants’ privacy?
5. What special issues are presented by using deception in Internet research?

## Ethical Guidelines, Your Research, and the Institutional Review Board

Now that you are familiar with the ethical principles for research with human participants, can you now proceed with your research? In days gone by, you could have done just that. Currently, it is likely that you will be required to have your research reviewed by an **institutional review board (IRB)**. If you are affiliated with any institution that receives federal funding and your research does not fall into an exempted category, you must have your research screened for ethical treatment of participants before you can begin to conduct your research. The role of the IRB is to ensure that you adhere to established ethical guidelines.

Submitting your research to the IRB for review involves drafting a proposal. The form of that proposal varies from institution to institution. However, an IRB requires certain items of information to evaluate your proposal. Information will be needed concerning how participants will be acquired, procedures for obtaining informed consent, experimental procedures, potential risks to the participants, and plans for following up your research with reports to participants. Depending on the nature of your research, you may be required to submit a draft of an “informed-consent form” outlining to your participants the nature of the study. Additional sections would be added to the consent form if your research participants will be paid, may sustain injury, or will incur any additional expenses (e.g., transportation costs and research-related supplies). Each institution, however, may have additional requirements for what must be included in an informed-consent form. Additionally, requirements for informed-consent forms may change frequently within an institution. Before using any consent form, you should consult your IRB and ensure that your form complies with its requirements.

You may see these preliminary steps as unnecessary and, at times, a bother. After all, aren't you (the researcher) competent to determine whether participants are being treated ethically? Although you may be qualified to evaluate the ethics of your experiment, you still have a vested interest in your research. Such a vested interest may blind you to some ethical implications of your research.

The IRB is important because it allows a group of individuals who do not have a vested interest in the research to screen your study. The IRB review and approval provides protection for both you and the sponsoring institution. If you choose to ignore the recommendations of the IRB, you may be assuming legal liability for any harm that comes to people as a result of participation in your research. In the long run, the extra time and effort needed to prepare the IRB proposal is in the best interests of the sponsoring institution, the participant, and you.

One factor that both the IRB and the researcher must assess is the risk–benefit ratio of doing research. Research may involve some degree of risk to participants, ranging from minimal to very high. This risk might involve psychological and/or physical harm to the participants. For example, a participant in an experiment on the effects of stress on learning might be subjected to stimuli that create high-level stress. It is possible the participants might be harmed psychologically by such high levels of stress. The researcher and the IRB must determine if the benefits of the research (new techniques for handling stress discovered, new knowledge about the effects of stress, etc.) outweigh the potential risks to the participant. In the event that high-risk

research is approved by the IRB, the researcher will be required to take steps to deal with any harmful side effects of participation in such research. For example, you may have to provide participants with the opportunity to speak to a counselor if they have an adverse reaction to your study.

One final note on the role of the IRB is in order. Many researchers view the IRB as an annoyance and an impediment to their research (Fiske, 2009). However, an IRB serves an important function. It ensures that your research conforms to accepted ethical principles and protects you from liability in case a participant suffers harm in your study. Susan Fiske (2009) states that IRBs work well when they adhere to two principles. First, they must act to protect human research participants against harm and unethical treatment. Second, IRBs can also serve to promote research by adequately training IRB staff and researchers concerning the IRB's function. Improving communication between researchers and IRB members is also part of this second function. With improved communication the review process can be viewed more as a collaborative process than one where the IRB mandates certain rules and procedures.

## QUESTIONS TO PONDER

---

1. What role does an institutional review board (IRB) play in the research process?
2. Why is IRB review important?
3. What are the IRB's two roles?

## ETHICAL CONSIDERATIONS WHEN USING ANIMAL SUBJECTS

---

You might be thinking to yourself, at this point, that with all of the rules and regulations governing the use of human research participants, you will circumvent them by doing your research using animals. After all, animals aren't people and probably won't have the same restrictive ethical rules and guidelines applying to them. Think again! If you choose to use animals in your research, you will have to adhere to a set of ethical guidelines that are just as comprehensive as those covering research with humans.

It is certainly true that you can carry out experiments with animals that are not ethically permissible with human participants. For example, you may do physiological research on the brain that involves systematically destroying parts of the brain. Such research, of course, would not be possible with human participants. We doubt that anyone would willingly give informed consent to have parts of the brain destroyed in the name of science. However, such techniques can (and have) been used with animal subjects.

Does this mean that if you use animals in your research you have a free hand to do anything you please? The answer is no. If you use animals in research, you are bound by a code of ethics, just as when you use human participants. This ethical code

**TABLE 7-3 2002 APA Ethical Code for the Care and Use of Animal Subjects**

1. Psychologists acquire, care for, use, and dispose of animals in compliance with current federal, state, and local laws and regulations, and with professional standards.
2. Psychologists trained in research methods and experienced in the care of laboratory animals supervise all procedures involving animals and are responsible for ensuring appropriate consideration of their comfort, health, and humane treatment.
3. Psychologists ensure that all individuals under their supervision who are using animals have received instruction in research methods and in the care, maintenance, and handling of the species being used, to the extent appropriate to their role.
4. Psychologists make reasonable efforts to minimize the discomfort, infection, illness, and pain of animal subjects.
5. Psychologists use a procedure subjecting animals to pain, stress, or privation only when an alternative procedure is unavailable and the goal is justified by its prospective scientific, educational, or applied value.
6. Psychologists perform surgical procedures under appropriate anesthesia and follow techniques to avoid infection and minimize pain during and after surgery.
7. When it is appropriate that an animal's life be terminated, psychologists proceed rapidly, with an effort to minimize pain and in accordance with accepted procedures.

SOURCE: APA, 2002.

specifies how animals may be treated, housed, and disposed of after use (Table 7-3). The U.S. Public Health Service (2002) has endorsed a set of principles for the care and use of animals that is strikingly similar to the APA's ethical principles. (These principles can be found at <http://www.nal.usda.gov/awic/legislat/awa.htm>.)

These guidelines make it clear that if you use animals in your research you must follow all applicable laws and closely supervise all procedures involving animals, including procedures carried out by laboratory assistants. They also make clear your responsibility to minimize discomfort, illness, and pain of the animals and to use painful procedures only if alternatives are not available.

### The Institutional Animal Care and Use Committee

Just as proposals for research using human participants must be reviewed and approved by an IRB before the research can be conducted, so proposals for research using animal subjects must be reviewed and approved by an **institutional animal care and use committee (IACUC)**. According to the *Guide for the Care and Use of Laboratory Animals* (National Research Council, 1996), committee membership should include the following:

- ♦ A doctor of veterinary medicine, who is certified . . . or has training or experience in laboratory animal science and medicine or in the use of the species in question,
- ♦ At least one practicing scientist experienced in research involving animals,
- ♦ At least one public member to represent general community interests in the proper care and use of animals. Public members should not be laboratory animal users, be affiliated with the institution, or be members of the immediate family of a person who is affiliated with the institution.

In practice, such committees are usually larger than this minimum. In colleges and universities, it is common to find representatives from departments that do *not* use animals in their research or teaching, as well as from those that do, and at least one student representative. The Purdue Animal Care and Use Committee (PACUC) at Purdue University includes more than 30 members and has a full-time staff, including specialists in laboratory animal science and veterinary medicine.

The use of animal subjects for research or teaching is regulated by the federal government, which mandates oversight by the IACUC and by the U.S. Department of Agriculture, as well as by various state and local agencies. The strict requirements for institutional care and use of animals under federal jurisdiction are given in the *Guide for the Care and Use of Laboratory Animals* (National Research Council, 1996). (You can find this publication online at <http://books.nap.edu/readingroom/books/labrats/>.)

Before you begin conducting your research using animal subjects, you should familiarize yourself with the principles for the care and use of animals and design your research accordingly. Before you can begin testing, you must submit a research protocol to your IACUC, describing what animals you plan to use in your research, how you plan to use them, and justifying your decisions concerning the species and number of animals to be used and the specifics of your procedure. Only when your protocol has been formally approved by the IACUC will you be permitted to obtain your animals.

Finally, keep in mind that ethical treatment of animals is in your best interest as a researcher. Ample evidence shows that mistreatment of animals (such as rough handling or housing them under stressful conditions) leads to physiological changes (e.g., housing animals under crowded conditions leads to changes in the adrenal glands). These physiological changes may interact with your experimental manipulations, perhaps damaging the external validity of your results. Proper care and handling of your subjects helps you obtain reliable and generalizable results. Thus, it is to your benefit to treat animal subjects properly.

### Cost–Benefit Assessment: Should the Research Be Done?

Even though a study is designed to conform to ethical standards for the use of animal subjects—giving proper care and housing, avoiding unnecessary pain or hardship, and so on—this does not automatically mean that the study should be done. Your decision to go ahead with the study should be based on a critical evaluation of the cost of the study to the subjects weighed against its potential benefits, otherwise known as the *cost–benefit ratio*. Cost to the subjects includes such factors as the stressfulness

of the procedures and the likely degree of discomfort or suffering that the subjects may experience as a result of the study's procedures. The potential benefits of the study include the study's possible contribution to knowledge about the determinants of behavior, its ability to discriminate among competing theoretical views, or its possible applied value in the real world.

Conducting an unbiased evaluation is not easy. Having designed the study, you have a certain vested interest in carrying it out and must guard against this bias. Yet if you reject a study because its potential findings do not have obvious practical application, you may be tossing out research that would have provided key insights necessary for the development of such applications. The history of science is littered with research findings whose immense value was not recognized at the time they were announced.

Despite these difficulties, in most cases it is possible to come up with a reasonable assessment of the potential cost–benefit ratio of your study. For example, imagine you have designed a study to evaluate the antianxiety effect of a certain drug (paramethyldoubleloop). You have no particular reason to believe that it has any effect on anxiety; in fact, its chemical structure argues against such an effect. However, you have a sample of the drug, and you're curious. Your subjects (rats) will have to endure a procedure involving water deprivation and exposure to foot shock in order for you to assess the effect of the drug. Given that the cost in stress and discomfort to the rats is not balanced against any credible rationale for conducting the study, you should shelve the study.

### QUESTIONS TO PONDER

---

1. What are the ethical guidelines you must follow when using animal subjects?
2. What is the composition of the institutional animal care and use committee and why is review important?
3. How does a cost–benefit analysis enter into one's decision to conduct a study using animal subjects?

### TREATING SCIENCE ETHICALLY: THE IMPORTANCE OF RESEARCH INTEGRITY AND THE PROBLEM OF RESEARCH FRAUD

---

Thus far, we have made a case for you to treat your human participants or animal subjects ethically and in a manner consistent with all relevant professional and government regulations. However, your responsibility to be an ethical researcher does not stop with how you treat your participants or subjects. You also have an obligation to treat your science ethically and with integrity. This is stated clearly in Section C (Integrity) of the ethical code of the APA: Psychologists seek to promote accuracy, honesty, and truthfulness in the science, teaching, and practice of psychology. In these activities psychologists do not steal, cheat, or engage in fraud, subterfuge, or intentional misrepresentation of fact (APA, 2002).

This ethical principle should not be taken lightly. Fraudulent or otherwise dishonest research practices can erode the public's confidence in scientific findings. It can also lead to potentially harmful outcomes for large groups of people. For example, fraudulent breast cancer research done in the 1990s suggested that the less radical *lumpectomy* (where only the tumor and surrounding tissue are removed) was just as effective as the more radical *mastectomy* (where an entire breast and surrounding tissue are removed). It turned out that the researcher, Dr. Roger Poisson, a noted cancer researcher, admitted that he had falsified his data concerning clinical tests of the two surgical procedures. He had allowed women into his research who were in more advanced stages of cancer than were to be permitted in the study, and he had reported on the progress of women who had died. He kept two sets of files on his research, one false and one truthful.

As a result of Poisson's unethical conduct, confidence in the lumpectomy versus mastectomy research was shaken. It also called into question the honesty of the entire scientific community. The public could no longer be sure that the results coming out of research laboratories could be trusted.

The preceding example illustrates how the process of science can be subverted by a dishonest scientist. Expectations of a researcher also can affect the outcome of a study. The case of Poisson and the impact of researcher expectations reveal an important truth about research in the social and behavioral sciences: It is a very human affair. The research process benefits from all the good qualities of human researchers: ingenuity, dedication, hard work, a desire to discover the truth, and so on. However, as in any human endeavor, the more negative qualities of human researchers also may creep into the research process: ambition, self-promotion, ego, securing and keeping a job, and obtaining scarce grant money. Donning a lab coat does not guarantee that a person checks his or her ambitions, flaws, desires, and needs at the laboratory door (Broad & Wade, 1983).

Research fraud can have direct financial and medical effects on participants in that research (Barrett & Jay, 2005). For example, in one case, a depressed patient was told by his doctor that there was a new drug to treat depression. The patient was told that because the drug was not approved in the United States he would have to sign a receipt for the drug. Upon investigation it was found that the patient had actually signed a consent form to participate in research using the unlicensed drug. His continued depression concerned his parents, who took him for more conventional treatment (Barrett & Jay, 2005).

Don't get the idea that fraud is a problem only in medical research. The U.S. **Office of Research Integrity (ORI)** is an office within the U.S. Department of Health and Human Services that oversees the integrity of the research process. The ORI documents and investigates cases of research fraud in science, including psychological research. For example, a 2006 case involving a former psychology graduate student at UCLA found that she had "engaged in scientific misconduct by falsifying or fabricating data and statistical results for up to nine pilot studies on the impact of vulnerability on decision making from Fall 2000 to Winter 2002 as a basis for her doctoral thesis research" (U.S. Office of Research Integrity, 2006a). The falsified data were used in a manuscript submitted to the journal *Psychological Science* and in a National Institutes of Mental Health grant proposal.

## QUESTIONS TO PONDER

---

1. What does the APA ethical code say about research integrity?
2. Why should we be concerned with research fraud?
3. What is the ORI and what does it do?

### What Constitutes Fraud in Research?

The ORI (2007, p. 2) defines three categories of research fraud:

1. Data fabrication: Making up data or results and reporting on them.
2. Falsification: Manipulating research materials, equipment, or processes, or changing or omitting data or results such that the research is not accurately represented in the research record.
3. Plagiarism: The appropriation of another person's ideas, processes, results, or words without giving appropriate credit.

According to the ORI (2007), honest errors and differences of scientific opinion do not constitute research fraud.

Perhaps the most harmful, but rare, form of research fraud is the outright fabrication of data (Broad & Wade, 1983). A scientist may fabricate an entire set of data based on an experiment that might never have been run or replace actual data with false data. Other forms of fraud in research include altering data to make them “look better” or fit with a theory, selecting only the best data for publication, and publishing stolen or plagiarized work (Broad & Wade, 1983). Altering or otherwise manipulating data in order to achieve statistical significance (e.g., selectively dropping data from an analysis) also would constitute research fraud.

Broad and Wade (1983) also suggest that using the *least-publishable-unit rule*, which involves getting several small publications out of a single experiment (as opposed to publishing one large paper), might be considered dishonest. Research fraud can occur if scientists sabotage each other's work. Claiming credit for work done by others also could be considered fraud. If, for example, a student conceptualizes, designs, and carries out a study but a professor takes senior author status on the publication, this would be considered fraud. It is also dishonest to attach your name to research that you had little to do with, just to pad your résumé. Some articles may have as many as 10 or more authors. Each of the junior authors may have had some minor input (such as suggesting that Wistar rats be used rather than Long–Evans rats). However, that minor input may not warrant authorship credit. Finally, plagiarism, in which a researcher uses another person's work or ideas without proper acknowledgment, is also a form of research fraud.

### The Prevalence of Research Fraud

At one time, the editor of *Science* stated that 99.9999% of scientific papers are truthful and accurate (Bell, 1992). The U.S. Office of Research Integrity (2007) found fraud in 10 out of 28 cases that it closed (6 cases of data falsification and 4 cases of



falsification and fabrication). A survey by Geggie (2001) of medical consultants in England found that 55.7% of respondents reported witnessing some form of research misconduct firsthand. Additionally, 5.7% reported engaging in misconduct themselves, and 18% indicated that they would consider engaging in misconduct in the future or were unsure about whether they would engage in research misconduct.

So, the numbers don't seem to be huge. However, despite this optimism, critics suggest that it is not possible to exactly quantify research fraud (Bell, 1992). For one thing, fraud may not be reported, even if it is detected. In Poisson's case, for example, some evidence exists that there was a suspicion of fraud as early as 1990. In addition, in one survey (cited in Bell, 1992), many researchers who suspected that a colleague was falsifying data did not report it. Fraud may also go unreported because the liabilities associated with "blowing the whistle" can be quite severe. Whistle-blowers may be vilified, their credibility is called into question, and they may, perhaps, even be fired for "doing the right thing." Thus, the relatively few celebrated cases of fraud reported may be only the tip of the iceberg.

Regardless of how low the actual rate of fraud in science turns out to be, even a few cases can have a damaging effect on the credibility of science and scientists (Broad & Wade, 1983). Erosion of the credibility of science undermines the public's confidence in the results that flow from scientific research. In the long run, this works to the detriment of individuals and of society.

### Explanations for Research Fraud

Why would a scientist perpetrate a fraud? There are many reasons. Fraud may be perpetrated for personal recognition. Publishing an article in a prestigious journal is a boost to one's self-esteem. Personal pressure for such self-esteem and recognition from others can motivate a person to falsify data or commit some other form of research fraud.

The pursuit of money is a major factor in fraudulent research (Bell, 1992). Doing research on a large scale takes quite a bit of money, and researchers are generally not wealthy and cannot fund their own research. Nor can financially strapped universities or hospitals provide the level of funding needed for many research projects. Consequently, researchers must look to funding agencies such as the National Science Foundation, the National Institute of Mental Health, or some other available funding source. The budgets for these agencies are typically limited with respect to the number of applications that can be accepted for funding. Consequently, competition for research funding becomes intense. In addition, it is generally easier to obtain grant money if you have a good track record of publications. The pressure for obtaining scarce grant money can lead a person to falsify data in order to "produce" and be in a good position to get more funding. Moreover, at some universities, obtaining grants is used as an index of one's worth and may even be a requirement of retaining one's job. This can add additional pressure toward committing research fraud.

Another reason for fraud in research relates to the tenure process within the academic environment. A new faculty member usually has 5 years to "prove" him- or herself. During this 5-year probationary period, the faculty member is expected to publish some given quantity of research articles. At high-power, research-oriented universities, the number of publications required may be large, creating a strong

“publish or perish” atmosphere. This atmosphere seems to have grown stronger over the past 40 years. When James D. Watson (Nobel Prize winner with Francis Crick for his discovery of the DNA double helix) was a candidate for tenure and promotion at Harvard University in 1958, he had 18 publications. By 1982, 50 publications were required for the same promotion (Broad & Wade, 1983). This need to publish as many papers as possible in a relatively short period of time can lead to outright fraud and/or *vita* (résumé) padding using the least-publishable-unit rule. Finally, fraud in research can arise from scientific “elitism” (Broad & Wade, 1983). Sometimes we see fraud committed by some of the biggest names in science because their elite standing in the scientific community shields their work from careful scrutiny.

### Dealing With Research Fraud

Bell (1992) points out that science has three general methods for guarding against research fraud: the grant-review process, the peer-review process for publication, and replication of results. Bell points out that, unfortunately, none of these is effective in detecting fraud. Editors may be disinclined to publish papers that are critical of other researchers, let alone that make accusations of fraud (Bell, 1992). In addition, replication of experiments is expensive and time consuming and unlikely to occur across labs (Bell, 1992). Even if a finding cannot be replicated, that does not necessarily mean fraud has occurred.

One way to deal with research fraud is to train students in the ethics of the research process. Students should learn, early in their academic careers, that ethical research practice requires scientific integrity and that research fraud is unethical. Unfortunately, students are often not taught this lesson very well. Michael Kalichman and Paul Friedman (1992) conducted a survey of biomedical science trainees and found that only 24% indicated that they had received training in scientific ethics. Additionally, 15% said that they would be willing to alter or manipulate data in order to get a grant or a publication. Geggie (2001) found that only 17% of respondents had received any training in research ethics. A study reported by David Wright, Sandra Titus, and Jered Cornelison (2008) paints an even bleaker picture. These researchers reviewed cases between 1990 and 2004 in which the ORI had found research fraud by research trainees (graduate students, lab assistants, and postdoctoral fellows) to see if their mentors had monitored their work. Wright et al. found that the mentors failed to review raw data in 73% of the cases and set no standards in 62% of the cases. In 53% of the cases the relationship between the mentors and trainees was stressful. Mentors were doing very little to reduce the likelihood of research fraud.

The ORI has a program designed to educate scientists about research fraud. One such program is the Responsible Conduct of Research (RCR) program. This program includes educational experiences centering on issues such as research misconduct, responsible authorship, and peer review. Educational materials for this program can be found at <http://ori.dhhs.gov/education/products/>. Another ORI education effort is an RCR exposition. At the exposition, various “vendors” can showcase their programs and products designed to reduce research fraud.

Jane Steinberg (2002) indicates that another safeguard against fraud is to make it clear to scientists and assistants that they will be caught if they commit scientific fraud. Steinberg suggests that researchers check data often and openly in front of

those who collect and analyze the data. Questions should be asked about any suspicious marks on datasheets or changes/corrections made to the data.

Probably the best guard against fraud in science is to imbue researchers during the training process with the idea that ethical research means being honest. This process should begin as early as possible. Steinberg (2002) suggests that teaching about research fraud should begin in psychology students' research methods courses. She recommends that students be presented with cases of research fraud. Those cases should be discussed and evaluated carefully. Students should learn the implications of research fraud for researchers themselves, their field of study, and the credibility of science (Steinberg, 2002). The short-term consequences (loss of a job, expulsion from school, etc.) and long-term consequences (harm to innocent individuals because of false results, damage to the credibility of science, etc.) should be communicated clearly to researchers during their education and training.

Another strategy suggested by Steinberg (2002) is to contact research participants after they have participated in a study to see if they actually participated. Participants should be asked if they actually met with the person running the study, whether they met eligibility requirements, if they knew the person running the study beforehand, and if the study ran for the appropriate amount of time. Similar steps can be taken with animal subjects by carefully scrutinizing animal use records and laboratory notes (Steinberg, 2002).

When fraud does occur, scientists should be encouraged to blow the whistle when they have strong proof that fraud took place. The U.S. Office of Research Integrity (2009) suggests that whistle-blowers are a crucial component in the fight against fraud in science. The ORI recommends that, before making an allegation of research fraud, the whistle-blower familiarize him- or herself with the policies of the institution, find out what to include in a report, and find out to whom the report should be given. The whistle-blower also should find out about protection against retaliation and about the role that he or she will play after the report is made. The ORI underscores the need for institutions to protect whistle-blowers from negative consequences.

A survey commissioned by the ORI (1995) found that 30.9% of whistle-blowers studied reported no negative consequences for their actions. However, 27.9% reported at least one negative consequence, and 41.2% reported multiple negative outcomes. Those negative outcomes included being pressured to drop the charges (42.6%), being hit with a countercomplaint (40%), being ostracized by coworkers (25%), or being fired or not receiving tenure (23.6%). Thus, the climate for whistle-blowers is quite hostile. For example, Stephen Bruening based a recommendation that retarded children be treated with stimulants (most research and practice suggested using tranquilizers) on years of fraudulent data. Robert Sprague exposed Bruening's fraud and was subjected to pressure from members of the University of Pittsburgh administration not to pursue his allegations against Bruening. Sprague was even threatened with a lawsuit.

Finally, a researcher must determine whether fraud has actually occurred. In some cases, this may be relatively easy. If a scientist knows for a fact that a particular study reported in a journal was never done, fraud can be alleged with confidence. In other cases, fraud may be detected by noticing strange patterns in the data reported. This is essentially what happened in the case of Cyril Burt, whose research had found strong correlations between the intelligence test scores of identical twins. Some researchers

noted that some of Burt's correlations remained invariant from study to study even though the numbers of participants on which the correlations were based changed. This, and the fact that an assistant whom Burt claimed helped him could not be found, served as the foundation of what seemed like a strong case against Burt based on circumstantial evidence.

Burt's posthumous reputation has been ruined and his work discredited. However, Joynson (1989) has reevaluated the Burt case and has provided convincing alternative explanations for the oddities in Burt's data. Joynson maintains that Burt did not deliberately perpetrate a fraud on science and that Burt's name should be cleared. At this point, the jury is still out on Burt's conduct as to whether he committed outright fraud. Even if Burt did not commit fraud, he was willing to misrepresent his data and recycle old text (Butler & Petrulis, 1999). On the other hand, there are those who contend that the evidence shows that Burt was guilty of fraud beyond any reasonable doubt (Tucker, 1997).

### QUESTIONS TO PONDER

---

1. What constitutes research fraud, and why does it occur?
2. How prevalent is research fraud?
3. How can research fraud be dealt with?

### SUMMARY

---

After you have developed your research idea into a testable hypothesis and settled on a research design, your next step is to recruit participants or subjects for your study. Before you can proceed with your study, however, it must be evaluated for ethical issues. A review board will determine if your research protocol adheres to accepted ethical guidelines.

You must consider the ethics of your research when human participants are chosen for study. Concern over the ethical treatment of participants can be traced back to the Nuremberg trials after World War II. During those trials, medical experiments conducted on inmates in concentration camps came to light. Because of the treatment of individuals in those experiments, the Nuremberg Code was developed to govern experiments with humans. The Declaration of Helsinki expanded on the concepts embodied in the Nuremberg Code and specified a set of ethical principles governing medical research.

The Belmont Report defined three basic principles that apply to all research with human participants. *Respect for persons* states that research participants should be autonomous and allowed to make their own decisions and that participants with limited autonomy deserve special treatment. *Beneficence* states that research participants must have their well-being protected. Beneficence embodies two elements: do no harm and maximize benefits while minimizing harm. *Justice* divides the burdens and benefits equally between the researcher and participant. Many of the ethical rules and guidelines that researchers follow flow from these three principles.

The APA developed a code of ethics for treatment of human participants in research that is based on the Nuremberg Code. This is the Ethical Principles of Psychologists and Code of Conduct 2002. Ethical treatment of participants in an experiment requires voluntary participation, informed consent, the right to withdraw, the right to obtain results, and the right to confidentiality (among others).

Because of continued concern over ethical treatment of human research participants and some high-profile cases of research that were ethically questionable, the U.S. Department of Health and Human Services issued its own set of guidelines for research using human participants. These guidelines apply to all research with human participants except for some research that meets certain criteria. The guidelines mandate committee review and approval of research and mandate special protections for vulnerable populations.

The Internet has provided a rich new venue for researchers. Some research falls easily under established ethical guidelines. However, other research (e.g., participant observation of chat rooms) poses special ethical questions. These special ethical questions fall into three areas: obtaining informed consent, maintaining privacy and confidentiality of participants, and using deception in Internet research.

The Institutional Review Board (IRB) is a committee that screens research proposals using humans as participants to ensure that the participants are treated ethically. When a research proposal is submitted to the IRB, it normally includes a description of how participants will be acquired, procedures for obtaining informed consent, experimental procedures, potential risks to the participants, and plans for following up your research with reports to participants. Depending on the nature of your research, you may be required to submit a draft of an informed-consent form outlining to your participants the nature of the study. IRB review is important because it allows a group of individuals with no vested interest in your research to ensure that ethical guidelines are followed.

A large amount of psychological research uses animal subjects. Animals are preferred to humans in situations in which experimental manipulations are unethical for use with humans. However, if you use animal subjects, you are still bound by an ethical code. Animals must be treated humanely. Any research proposing to use animals as subjects must be reviewed by an institutional animal care and use committee (IACUC). The IACUC includes, among others, a veterinarian, a scientist experienced in animal research, and an interested member of the public. There are also federal, state, and local regulations that govern the use of animals in research that must be followed. It is to your advantage to treat your animals ethically because research shows that mistreated animals may yield data that are invalid. Even if your proposal for animal research meets ethical requirements, you still must do a cost-benefit analysis to determine if the study is worth doing.

In addition to treating human participants and animal subjects ethically, you are obligated to treat your science ethically. This means that you should “seek to promote accuracy, honesty, and truthfulness in the science, teaching, and practice of psychology.” This admonition from the APA should be taken seriously. Fraudulent, dishonest research has the potential to harm research participants and the credibility of scientists and science in general. Fraud in science is a problem that damages the credibility of science and its findings. Although it is rare, fraud does occur. Fraud

includes outright fabrication of data, altering data to look better, selecting only the best data for publication, using the least publishable rule, and taking credit for another's work. Motivation to commit fraud may stem from the desire to publish in prestigious journals, pressure to obtain scarce research funding, pressure to obtain publications necessary for tenure, and scientific elitism. The best way to deal with fraud in research is to train scientists so that they understand the importance of honesty in research.

### KEY TERMS

---

informed consent

Nuremberg Code

Declaration of Helsinki

Belmont Report

respect for persons

beneficence

justice

Ethical Principles of Psychologists and Code of Conduct 2002

institutional review board (IRB)

institutional animal care and use committee (IACUC)

Office of Research Integrity (ORI)

# Using Nonexperimental Research

In Chapter 4, we distinguished between correlational (nonexperimental) research (which involves observing variables as they exist in nature) and experimental research (which involves manipulating variables and observing how those manipulations affect other variables). In this chapter, we introduce you to several nonexperimental (correlational) research designs and to observational techniques often associated with them. As you read about the observational techniques, bear in mind that many of them also can be used when conducting experimental research.

## CONDUCTING OBSERVATIONAL RESEARCH

Although all research is observational (in the sense that variables are observed and recorded), the observational research designs described in this chapter are purely observational in two senses: (1) They are correlational designs and thus do not involve manipulating independent variables, and (2) all use trained researchers to observe subjects' behaviors. This section describes how to make and assess behavioral observations. Before we look at the “nuts and bolts” of observational research, let's take a look at an example of observational research.

### An Example of Observational Research: Are Children Really Cruel?

It is often said that children can be cruel. Children often tease or socially exclude other children who don't fit in with the peer group. Is it true that children, given the opportunity, will be “cruel” to another child? Will children display aggression or social exclusion against another child who doesn't fit in? An observational study by Marion Underwood, Betrina Scott, Mikal Galperin, Gretchen Bjornstad, and Alicia Sexton (2004) sought to find out.

## CHAPTER OUTLINE

### Conducting Observational Research

An Example of Observational Research: Are Children Really Cruel?

Developing Behavioral Categories

Quantifying Behavior in an Observational Study

Recording Single Events or Behavior Sequences

Coping With Complexity

Establishing the Reliability of Your Observations

Sources of Bias in Observational Research

Quantitative and Qualitative Approaches to Data Collection

### Nonexperimental Research Designs

Naturalistic Observation

Ethnography

Sociometry

The Case History

Archival Research

Content Analysis

### Meta-Analysis: A Tool for Comparing Results Across Studies

Step 1: Identifying Relevant Variables

Step 2: Locating Relevant Research to Review

Step 3: Conducting the Meta-Analysis

Drawbacks to Meta-Analysis

### Summary

### Key Terms

Participants in this study were pairs of children who were close friends. Both male and female friend pairs were included in the study. Children from three grade levels in school were included: those who had just completed fourth, sixth, and eighth grade. In the study, the friend pairs were told that they would play the game *Pictionary* with another child whom neither knew. The third child was actually an actor or actress working for the research team (more on this later). The game-playing session was conducted through four phases. In Phase 1, the two friends played the game while the third child was out of the room (ostensibly to finish filling out a questionnaire). In Phase 2, the actor rejoined the group and behaved in a friendly, neutral way. In Phase 3, the actor began verbally provoking the two friends. In Phase 4, the actor said that he or she had to go to the bathroom and left the friends alone for 2 minutes.

Throughout the session, the behavior of the two friends and the actor was recorded on videotape via four cameras mounted on the walls of the room, six feet above the floor. The cameras were covered with a plastic shield and were controlled remotely from another room. Close-up recordings were made of each child and of the group as a whole. Later, the videotapes were viewed by observers who coded the children's behavior along three behavioral dimensions: verbal social exclusion, verbal aggression, and verbal assertion. Underwood et al. (2004) found that both boys and girls used verbal social exclusion at about the same rates but in different situations. Boys were more socially exclusive when the child actor provoked the friends and was present. When the actor was out of the room, boys and girls used verbal social exclusion at about the same rates. Underwood et al. (2004) also found that fourth graders used more verbal social exclusion than eighth graders.

Now that you have seen how an observational study works, we can turn to the mechanics of performing observational research. The first step is to develop behavioral categories.

### Developing Behavioral Categories

**Behavioral categories** (also referred to as *coding schemes* or in animal research as *ethograms*) include both the general and specific classes of behavior that you are interested in observing. Each category must be operationally defined. For example, Underwood et al. (2004) defined the behavioral categories for their study of social exclusion as follows:

*Verbal social exclusion*: “Gossiping, planning to exclude the peer, emphasizing the friendship and the peer’s outsider status, and whispering” (p. 1545).

*Verbal aggression*: “Mockery, sarcasm and openly critical comments” (p. 1545).

*Verbal assertion*: “Saying ‘shhh!’ to the actor, telling the actor to stop cheating or to stop bragging, or disputing the actor’s comments” (p. 1545).

Developing behavioral categories can be a simple or formidable task. Recording physical characteristics of the subject is a relatively simple affair. However, when recording social behaviors, defining behavioral categories becomes more difficult. This is because coding socially based behaviors may involve cultural traditions that are not agreed on (e.g., coding certain speech as “obscene”) (Bakeman & Gottman, 1997).



Your behavioral categories operationally define what behaviors are recorded during observation periods, so it is important to define your categories clearly. Your observers should not be left wondering what category a particular behavior falls into. Ill-defined and ambiguous categories lead to recording errors and results that are difficult to interpret.

To develop clear, well-defined categories, begin with a clear idea about the goals of your study. Clearly defined hypotheses help narrow your behavioral categories to those that are central to your research questions. Also, keep your behavioral categories as simple as possible (Bakeman & Gottman, 1997) and stay focused on your research objectives. Avoid the temptation to accomplish too much within a single study.

One way to develop behavioral categories is to make informal, preliminary observations of your subjects under the conditions that will prevail during your study. During these preliminary observation periods, become familiar with the behaviors exhibited by your subjects and construct as complete a list of them as you can. Later, you can condense these behaviors into fewer categories, if necessary. Another way to develop behavioral categories is to conduct a literature search to determine how other researchers in your field define behavioral categories in research situations similar to your own. You might even find an article in which the researchers used categories that are nearly perfect for your study. Adapting someone else's categories for your own use is an acceptable practice. In fact, standardizing on categories used in previous research will enhance the comparability of your data with data previously reported.

Even if you do find an article with what appear to be the "perfect" categories, make some preliminary observations to be sure the categories fit your research needs. Take the time necessary to develop your categories carefully. In the long run, it is easier to adjust things before you begin your study than to worry about how to analyze data that were collected using poorly defined categories.

### Quantifying Behavior in an Observational Study

As with any other type of measure, direct behavioral observation requires that you develop ways to quantify the behaviors under observation. Methods used to quantify behavior in observational studies include the frequency method, the duration method, and the intervals method (Badia & Runyon, 1982).

**Frequency Method** With the *frequency method*, you record the number of times that a particular behavior occurs within a time period. This number is the frequency of the behavior. For example, Underwood et al. (2004) counted the number of statements made by children that were socially exclusive, aggressive, or assertive.

**Duration Method** With the *duration method*, your interest is in how long a particular behavior lasts. For example, you could record the duration of each verbally aggressive act displayed by children during a game-playing session. You can use the duration method along with the frequency method. In this case, you record both the frequency of occurrence (e.g., number of verbally aggressive acts) of a behavior and its duration (e.g., how long a verbally aggressive act lasts).

**Intervals Method** With the *intervals method*, you divide your observation period into discrete time intervals and then record whether a behavior occurs within each interval. For example, you might record whether an act of verbal exclusion occurs during successive 2-minute time periods. Ideally, your intervals should be short enough that only one instance of a behavior can occur during an interval. This was the method used by Underwood et al. (2004) in their social exclusion study. They divided observation periods into 10-second intervals and coded verbal exclusion, aggression, and assertiveness within those intervals.

### Recording Single Events or Behavior Sequences

Researchers doing observational studies have long recorded single events occurring within some identifiable observation period. Bakeman and Gottman (1997) advocate looking at *behavior sequences* rather than at isolated behavioral events. As an example, consider an observational study of language development in which you record the number of times that a parent uses language to correct a child's behavior. Although such data may be informative, a better strategy might be to record those same behaviors sequentially, noting which instances of language use normally follow one another. For example, is a harsh reprimand more likely to follow destructive behavior than nondestructive behavior? Recording such behavior sequences provides a more complete picture of complex social behaviors and the transitions between them.

Although recording behavior sequences requires more effort than recording single events, the richness of the resulting data may be well worth the effort. You can find more information about this method in Bakeman and Gottman (1997), *Observing Interaction: An Introduction to Sequential Analysis*. For a more advanced treatment, see Gottman and Roy (2008).

### Coping With Complexity

When you have defined your behavioral categories and settled on a method of quantifying behavior, you next must decide how to make your observations. Defining discrete time intervals during which to record behavior is easy enough, but actually recording the observations may be another matter. Take the example of observing the free-play behavior of preschool children.

Assume that you have clearly defined your behavioral categories and have decided to use the frequency method to quantify behavior. On Monday at 8 A.M., you arrive at the preschool classroom at which you intend to make your observations. Fourteen children are in the class. You sit in an observation room equipped with a one-way mirror and begin to observe the children in the classroom on the other side of the mirror. It doesn't take you long to realize that something is wrong. Your participants are running around in small groups, scurrying hither and yon. You cannot possibly observe all the children at once. Dejectedly, you leave the preschool and return home to try to work out an effective observation strategy.

This vignette illustrates an important fact about behavioral observation: Having clearly defined behavioral categories and adequate quantification methods does not guarantee that your observational techniques will work. Naturally occurring

behavior is often complex and fast paced. To make effective observations, you may need to use special techniques to deal with the rate at which the behaviors you wish to observe occur.

One solution to the problem is to sample the behaviors under observation rather than attempt to record every occurrence. Three sampling techniques from which to choose are time sampling, individual sampling, and event sampling (Conrad & Maul, 1981). Recording devices are also useful for observing behavior.

**Time Sampling** With *time sampling*, you scan the group for a specific period of time (e.g., 30 seconds) and then record the observed behaviors for the next period (e.g., another 30 seconds). You alternate between periods of observation and recording as long as necessary. Time sampling is most appropriate when behavior occurs continuously rather than in short bursts spaced over time and when you are observing large groups of subjects engaged in complex interactions.

**Individual Sampling** With *individual sampling*, you select a single subject for observation over a given time period (e.g., 10 minutes) and record his or her behavior. Over successive time periods, repeat your observations for the other individuals in the observed group. Individual sampling is most appropriate when you want to preserve the organization of an individual's behavior over time rather than simply noting how often particular behaviors occur.

**Event Sampling** In *event sampling*, you observe only one behavior (e.g., sharing behavior) and record all instances of that behavior. Event sampling is most useful when you can clearly define one behavior as more important than others and focus on that one behavior.

**Recording** You also could use recording devices to make a permanent record of behavior for later analysis as Underwood et al. (2004) did in their study. They installed four video cameras in the room unfamiliar to the children to record their behavior on tape.

Recording equipment has several advantages. First, because you have a permanent record, you can review your subjects' behavior several times, perhaps picking up nuances you might have missed in a single, live observation. Second, you can have multiple observers watch the recorded video independently and then compare their evaluations of behavior. (Although you can use multiple observers for live observations, it may be disruptive to your subjects to have several observers watching.) Finally, you may be able to hide a camera more easily than you can hide yourself. The hidden camera may be less disruptive to your subjects' behavior than an observer. This was the strategy used by Underwood et al. (2004). Recall that they had their cameras mounted on the walls of the room, covered with plastic shields and remotely controlled.

Making video recordings of behavior does not eliminate the need to classify the behaviors and to measure such aspects of the behaviors as frequencies and durations. Whether you perform these activities live or work from a recording, you will need a system for coding these characteristics.

One option is to develop a paper-and-pencil coding form similar to the one shown in Figure 8-1. Your observers would use the form to record the behaviors they see. Another option is to have observers speak into a handheld audio recorder.

Which of these two options you should choose depends on the nature of your study and limitations inherent in the situation. You can use paper-and-pencil coding sheets in just about any situation. They are quiet and, if properly constructed, efficient. They do have a few drawbacks, however. If you are requiring your observers to make extensive notes (not just checking behavioral categories), the task may become too complex and time consuming, especially if behaviors occur in rapid succession. In such cases, you might consider having your observers use audio recorders instead

Minutes	10-second intervals	Aggressive play	Nonaggressive play	Prosocial behavior	Solitary play	Verbal behavior	Talking with teacher
1	10						
	20						
	30						
	40						
	50						
	60						
2	10						
	20						
	30						
	40						
	50						
	60						
3	10						
	20						
	30						
	40						
	50						
	60						
10	10						
	20						
	30						
	40						
	50						
	60						

Subject \_\_\_\_\_

Observer \_\_\_\_\_

Date \_\_\_\_\_

Day care \_\_\_\_\_

FIGURE 8-1 Example of a paper-and-pencil coding sheet for an observational study.

of paper-and-pencil coding forms. The main advantage with this technique is that your observers will probably be able to speak into the recorder faster than they could make written notes. They also can keep their eyes on the subjects *while* making their notes. A disadvantage is that observers speaking into a recorder may disturb your subjects. Consequently, use this technique only when your observers are out of earshot of your subjects.

## QUESTIONS TO PONDER

---

1. What are the defining characteristics of observational research?
2. How are behavioral categories that are used in observational research developed?
3. What are the techniques used to make behavioral observations in observational research?
4. What is the distinction between recording single acts and behavior sequences?
5. What are the sampling techniques used to handle complexity when making behavioral observations?

## Establishing the Reliability of Your Observations

Assume that by now you have adequately defined the behavior you want to observe, developed a coding sheet, and worked out how you are going to observe behavior. You go into the field and begin making your observations. You come back with reams of data-coding sheets in hand and begin to summarize and interpret your data. You have apparently covered every possible base and believe that your observations accurately portray the observed behavior. But do they? Your observations may not be as accurate as you think. Your personal biases and expectations may have affected how you recorded the behavior observed. As with any measurement technique, when you conduct direct behavioral observations, you should make an effort to establish the reliability of your observations.

If you were the only observer, you could not firmly establish the reliability of your observations. To avoid the problem of single-observer idiosyncrasies, you should use multiple observers. This practice is generally preferred over single-observer methods.

When using multiple observers, you face the possibility that your observers will not agree when coding behavior. Theoretically, if you use well-trained observers and well-defined behavior categories, there should be a minimum of disagreement. However, disagreement is likely to arise despite your best efforts. Observers invariably differ in how they see and interpret behavior. Something as simple as a different angle of view can cause a disagreement. Disagreement also may arise if you have not clearly defined your behavioral categories. Because disagreement is likely to occur to one degree or another, you must assess **interrater reliability**, which provides an empirical index of observer agreement.

Bakeman and Gottman (1997) point out that there are three reasons to check for interrater reliability. First, establishing interrater reliability helps ensure that your observers are accurate and that you can reproduce your procedures. Second, you can check to see that your observers meet some standard that you have established. Third, you can detect and correct any problems with additional observer training.

There are several ways you can evaluate interrater reliability. In the sections that follow, we explore some of them.

**Percent Agreement** The simplest way to assess interrater reliability is to evaluate *percent agreement*. This method involves counting the number of times your observers agreed and dividing this number by the total number of observations. Specifically, you calculate percent agreement according to the following formula:

$$\frac{\text{Total number of agreements}}{\text{Total number of observations}} \times 100$$

For example, if your observers agreed on 8 of 10 observations, then the percent agreement would be

$$\frac{8}{10} \times 100 = 80\%$$

Of course, you want your percent agreement to be as high as possible, approaching 100%. However, for most applications, a percent agreement around 70% is acceptable.

Although percent agreement is a simple way to assess interrater reliability, the technique has drawbacks. First, if you define agreement as an exact match between observations, then percent agreement *underestimates* interrater agreement (Mitchell, 1979). You can reduce this problem somewhat by using a looser definition of *agreement*. Second, percent agreement gives you only a raw estimate of agreement. Some agreement between observers is to be expected based on chance alone. Percent agreement makes no provision for estimating the extent to which the agreement observed may have occurred by chance (Mitchell, 1979). Third, behaviors that occur with very high or low frequency may have extremely high levels of chance agreement. In those cases, percent agreement *overestimates* interrater agreement (Mitchell, 1979).

**Cohen's Kappa** A more popular method of assessing interrater reliability than percent agreement is **Cohen's Kappa**. Unlike percent agreement, Cohen's Kappa ( $K$ ) assesses the amount of agreement actually observed relative to the amount of agreement that would be expected by chance (via a statistical test; see Bakeman & Gottman, 1997). To use this method, you need to determine (1) the proportion of actual agreement between observers (actual agreement) and (2) the proportion of agreement you would expect by chance (expected agreement). Use these two values in the following formula (Bakeman & Gottman, 1997):

$$K = \frac{P_o - P_c}{1 - P_c}$$

where  $P_o$  is the observed proportion of actual agreement and  $P_c$  is the proportion of expected agreement.

Suppose you conducted a study of the relationship between the number of hours that an infant spends in day care and later attachment security. As your measure of attachment security, you have two observers watch a mother and her child for a 20-minute period. The coding scheme here is simple. All your observers are required to do is code the child's behavior as indicative of either a "secure attachment" or an "insecure attachment" within each of 20 one-minute observation periods. Sample coding sheets are shown in Figure 8-2. A mark in a cell indicates that the behavior of the child fell into that category.

The first step in computing Cohen's Kappa is to tabulate in a *confusion matrix* the frequencies of agreements and disagreements between observers (Bakeman & Gottman, 1997), as shown in Figure 8-3. The numbers on the diagonal (colored line in the figure) represent agreements, and the numbers off the diagonal represent disagreements. The numbers along the right edge and bottom of the matrix represent the row and column totals and the total number of observations.

		Observer 1		Observer 2	
		Secure	Insecure	Secure	Insecure
Observation period	1	/		/	
	2	/		/	
	3	/		/	
	4	/		/	
	5		/		/
	6	/		/	
	7	/			/
	8		/		/
	9		/		/
	10	/		/	
	11	/		/	
	12	/		/	
	13	/		/	
	14	/		/	
	15	/		/	
	16	/		/	
	17	/		/	
	18	/		/	
	19	/		/	
	20	/		/	

**FIGURE 8-2** Sample coding sheets for two observers counting "secure" and "insecure" behavior instances.

**FIGURE 8-3** Sample confusion matrix.

		Observer 1		
		Secure	Insecure	
Observer 2	Secure	16	0	16
	Insecure	1	3	4
		17	3	20

← Diagonal (agreement)

The next step is compute the value of Cohen's Kappa. First, determine the proportion of actual agreement by summing the values along the diagonal and dividing by the total number of observations:

$$P_o = \frac{16 + 3}{20} = .95$$

Next, find the proportion of expected agreement by multiplying corresponding row and column totals and dividing by the number of observations squared (Bakeman & Gottman, 1997):

$$P_c = \frac{(17 \times 16) + (3 \times 4)}{20^2} = .71$$

Finally, enter these numbers into the formula for Cohen's Kappa:

$$K = \frac{.95 - .71}{1 - .71} = .83$$

At this point, you have computed a reliability score of .83. What does this number mean? Is this good or bad? According to Bakeman and Gottman (1997), any value of .70 or greater indicates acceptable reliability.

**Pearson's Product-Moment Correlation** Pearson's product-moment correlation coefficient, or Pearson  $r$  (see Chapter 13), provides a convenient alternative to Cohen's Kappa for measuring interrater agreement. Table 8-1 shows the frequency of aggressive behavior among members of a hypothetical monkey colony over five 2-minute observation periods as coded by two observers. If your observers agree, Pearson  $r$  will be strong and positive. For example, Pearson  $r$  for the data shown in Table 8-1 is .90. This strong correlation (the maximum possible is 1.00) indicates substantial agreement. After calculating Pearson  $r$ , you can easily determine its statistical significance (see Chapter 13), an advantage over Cohen's Kappa.



**TABLE 8-1 Hypothetical Monkey Aggression Data Collected by Two Observers**

FREQUENCY OF AGGRESSIVE BEHAVIOR		
<i>Observation Period</i>	<i>Observer 1</i>	<i>Observer 2</i>
One	6	7
Two	2	4
Three	1	0
Four	5	7
Five	3	2

You must be cautious when using Pearson  $r$  to assess interrater agreement. Two sets of numbers can be highly correlated even when observers disagree markedly. This situation occurs when the *magnitudes* of the recorded scores increase and decrease similarly across observations by the two observers but differ in absolute value. For example, assume Observer 1 recorded 1, 2, 3, 4, and 5 and Observer 2 recorded 6, 7, 8, 9, and 10 over the same intervals. These numbers are perfectly correlated ( $r = 1.00$ ), yet the two observers never agreed on the actual numbers to record. You can check for this problem by comparing the means and standard deviations of the two sets of scores. If they are similar and Pearson  $r$  is high, you can safely assume that your observers agree.

**Intraclass Correlation Coefficient (ICC)** You can use the **intraclass correlation coefficient** ( $r_I$ ) to assess reliability if your observations are scaled on an interval or ratio scale of measurement. For example, you could use the ICC if you had observers count the frequency of aggressive behavior, which is a ratio-scaled measure. The ICC uses an analysis-of-variance approach to assess reliability. To compute an ICC, first construct a two-way table in which the columns represent observers' ratings (one column for each observer) and the rows represent participants (each row has the data from one participant), as shown in Table 8-2. The formula for calculating  $r_I$  uses the mean squares ( $MS$ ) components from the analysis of variance: the mean square within subjects ( $MS_W$ ) and mean square between subjects ( $MS_B$ ). The formula used to calculate  $r_I$  suggested by Shrout and Fleiss (1979) is

$$r_I = \frac{MS_B - MS_W}{MS_B + (k - 1)(MS_W)}$$

where  $k$  is the number of raters. For the data shown in Table 8-2, we have

$$r_I = \frac{14.6056 - .3833}{14.6056 + (2 - 1)(.3833)} = .95$$

ICC analysis is a flexible, powerful tool for evaluating interrater reliability. See Shrout and Fleiss (1979) and McGraw and Wong (1996) for in-depth discussions of ICC.

**TABLE 8-2** Hypothetical Data for an Intraclass Correlation Coefficient to Test Interrater Reliability

PARTICIPANT	RATER 1	RATER 2
1	8	10
2	6	5
3	2	2
4	7	7
5	5	5
6	6	9
7	1	1
8	6	6
9	7	8
10	8	7

**Dealing With Data From Multiple Observers** When multiple observers disagree, what should you do? If you have a high level of agreement, you can average across observers. For example, in Table 8-1, you can average across observers within each observation period to get a mean, or  $M$  [for the first period,  $M (6 + 7)/2 = 6.5$ ], and then obtain an overall average across observation periods. This gives you the average aggression shown during the observation period.

Another common method is to have observers meet and resolve any discrepancies. This method is practical when you have recorded the behavior electronically and can review it. Yet another method is to designate one of the observers as the “main observer” and the other as the “secondary observer.” Make this designation before you begin your observations. The observations from the “main observer” then serve as the numbers used in any data analyses. You use the observations from the “secondary observer” to establish reliability.

### Sources of Bias in Observational Research

Because observational research is a human endeavor, a degree of bias may contaminate the observations. One source of bias that can easily be avoided is observer bias. *Observer bias* occurs when your observers know the goals of a study or the hypotheses you are testing and their observations are influenced by this information. For example, suppose you have hypothesized that males will show more interpersonal aggression than females and have told your observers of this hypothesis. Suppose your observers see a male child roughly take a toy away from another child and later see a female child do the same. Because of observer bias, your observers may code the male child’s behavior, but not the female child’s behavior, as aggressive. This is the same problem discussed in Chapter 5 as experimenter bias, and the solution

is the same: Use a *blind observer*. A blind observer is one who is unaware of the hypotheses under test. (For more details about blind techniques, see Chapter 5.)

Another source of bias in observational research arises when observers interpret what they see rather than simply record behavior. We have all seen nature specials on television in which a researcher is observing animals in the wild (e.g., chimpanzees). Too often, those researchers infer intentions behind the behaviors they observe. When one chimp prods another with a stick, for example, the researcher may record the behavior as a “playful, mischievous attack.”

The problem with such inferences is that we simply do not know whether they are correct. We tend to read into the behaviors of animals the motivations and emotions we ourselves would likely experience in similar situations. However, the animal’s motivations and emotions may in fact be very different from ours. Stick to what is immediately apparent from the observation. If you have preserved the actual behavior in your records rather than your interpretation of the behavior, you can always provide an interpretation later. If new evidence suggests a different interpretation, you will have the original behavioral observations available to reinterpret.

When your subjects are people, you should still do your best to record behaviors rather than your interpretations of those behaviors. Piaget showed that inferences concerning the motivation and knowledge of children are often wrong, and the same is probably true of inferences about adults. Once again, if your data preserve the behavior rather than your interpretations of the behavior, you can always reinterpret your data later if required by new evidence.

### Quantitative and Qualitative Approaches to Data Collection

When making your observations and recording behavior, you can use two approaches to recording behavior. Counting and otherwise quantifying behavior yield **quantitative data**, which are expressed numerically. The main advantage of quantitative data is that a wide range of statistical tests is available for analyzing these data. (However, not all research situations lend themselves to quantitative data collection.) Using a *quantitative approach* to study client reactions to a new form of psychotherapy, you might have therapy clients rate how they feel about their therapy on rating scales, or you might count the number of times that a certain thing was mentioned (such as the warmth of the experimenter). In both instances, your data will be numbers that can be mathematically manipulated and analyzed with available descriptive and inferential statistics.

In some instances, you might consider collecting qualitative data. **Qualitative data** consist of written records of observed behavior that you analyze qualitatively. No numbers are generated on rating scales nor are there counts of behavior. Using a *qualitative approach* to study client reactions to a new therapy technique, you could interview clients and then review the interview protocols to extract themes that emerged via the interviews (e.g., clients’ impressions of the language used during the therapy). Because the data are qualitative, you cannot apply standard descriptive and inferential statistics to your data. In fact, analysis of qualitative data poses special problems for researchers. Usually, there are large amounts of raw data to deal with, and you will need specialized computer programs to analyze qualitative verbal information.

Depending on your research situation, you may collect only quantitative data, only qualitative data, or, as many studies do, a combination of the two. In the sections that follow, which introduce various nonexperimental methods, we present a mix of examples illustrating both quantitative and qualitative approaches.

## QUESTIONS TO PONDER

---

1. Why should you evaluate interrater reliability?
2. What are the techniques used to evaluate interrater reliability, and when would each be used?
3. How do you deal with data from multiple observers?
4. What are the sources of bias in observational research, and how can the bias be reduced?
5. What is the difference between quantitative and qualitative data?
6. What are the problems inherent in collecting qualitative data?

## NONEXPERIMENTAL RESEARCH DESIGNS

---

Now that you know how to develop and use direct behavioral measures, it is time to become familiar with several nonexperimental approaches to data collection. Keep in mind that in each of these designs you can apply any of the aforementioned observational methods to collect your data.

### Naturalistic Observation

**Naturalistic observation** involves observing your subjects in their natural environments without making any attempt to control or manipulate variables. For example, you might observe chimpanzees in their African habitat, children in a day-care center, shoppers in a mall, or participants in a court proceeding. In all these cases, you would avoid making any changes in the situation that might affect the natural, ongoing behaviors of your subjects.

**Making Unobtrusive Observations** Although you may not intend it, the mere act of observing may disturb the behavior of your subjects. Such disturbances may reduce the internal or external validity of your observations. To prevent this difficulty, you should make *unobtrusive observations*, or observations that do not alter the natural behaviors of your subjects.

Putting this requirement into practice may involve the use of special equipment. When studying the nesting habits of a particular species of birds, for example, you may have to build a blind (an enclosure that shields you from the view of your subjects) from which to make your observations. When studying social interactions among preschool children in a day-care center, you may have to make your observations

from behind a one-way mirror. In either case, you want to prevent your subjects from knowing that they are being observed.

Unfortunately, it is not always possible to remain hidden. For example, you may need to be closer to your subjects than a blind or observation room allows. In such cases, a widely used technique is to habituate your subjects to your presence (a fancy way of saying “letting your subjects get used to you”) before you begin making your observations. Habituating subjects involves gradually introducing yourself to the environment of your subjects. Eventually, your subjects will view your presence as normal and ignore you. If you were interested in observing children in a day-care center, for example, you might begin by sitting quietly away from the children (perhaps in a far corner of the room) until the children no longer paid attention to you. Gradually, you would move closer to the children, allowing them to habituate to your presence at each step before moving closer.

Habituation may be necessary even if you are going to videotape behavior for later analysis. The presence of a television camera in a room will attract attention at first. Allowing your subjects to habituate to the camera before you begin your observations will help reduce the camera’s disruptive effects.

You also can make observations unobtrusively by abandoning direct observations of behavior in favor of *indirect measures*. For example, to study recycling behavior, you could look through recycling bins or trash to gauge the extent to which your participants recycle and determine the types of materials they recycle. In this case, participants do not know that their behavior is being observed (unless you get caught snooping through their trash and recycling material).

**Advantages and Disadvantages of Naturalistic Observation** Naturalistic observation gives you insight into how behavior occurs in the real world. The observations you make are not tainted by an artificial laboratory setting, and therefore you can be reasonably sure that your observations are representative of naturally occurring behavior. In other words, properly conducted naturalistic observation has extremely high external validity.

Because naturalistic observation allows you only to describe the observed behavior, you cannot use this technique to investigate the underlying *causes* of those behaviors. In addition, naturalistic observation can be time consuming and expensive. Unlike some types of observation in which subjects in effect record their own data, naturalistic observation requires you to be there, engaged in observation, during the entire data-collecting period, which may last hours, days, or longer. Also, getting to the natural habitat of your subjects may not be easy. In some cases (such as observing chimpanzee behavior in the wild, as Jane Goodall did), naturalistic observation requires traveling great distances to reach the habitat of your subjects.

**An Example of Naturalistic Observation: Communication Among the Elderly** As an example of naturalistic observation consider a study of communication patterns among elderly patients with aphasia (loss of speech functions) resulting from a stroke, conducted by Brownyn Davidson, Linda Worall, and Louise Hickson (2003). In this study, participant-observers made observations of communication during patients’ everyday activities. (We discuss participant observation below in the section on

ethnography.) Davidson et al. instructed the participant-observers not to initiate any communication with the patients but to make limited responses when addressed by patients. Observations were also made of healthy older adults.

Each patient was observed for 8 hours over three randomly determined time periods within a week. Davidson et al. (2003) had observers code a number of communication behaviors, including conversations, greetings, talking to pets, talking on the telephone, writing/word processing, and storytelling. Davidson et al. found that patients' conversations were most likely to occur at home or in some social group. They also found that aphasic and healthy elderly adults engaged in communication on a range of topics that were similar to those of healthy elderly adults. The main difference between the aphasic and elderly adults was that aphasics engaged in quantitatively less communication than healthy elderly adults. Finally, aphasics were far less likely to engage in storytelling than healthy elderly adults.

## Ethnography

In **ethnography** a researcher becomes immersed in the behavioral or social system being studied (Berg, 2009). The technique is used primarily to study and describe the functioning of cultures through a study of social interactions and expressions between people and groups (Berg, 2009). Like an investigative reporter or undercover police officer, you insinuate yourself within a group and study the social structures and interaction patterns of that group from within. Your role as a researcher is to make careful observations and record the social structure of the group that you are studying.

Ethnography is a time-tested research technique that has been popular, especially in the field of anthropology. However, it is also used in sociological and psychological studies of various behavior systems. For example, ethnography is used to study client and therapist perceptions of marital therapy (Smith, Sells, & Clevenger, 1994), the learning of aggressive behaviors in different cultures (e.g., Fry, 1992), the assassination of John F. Kennedy (Trujillo, 1993), the sudden death of loved ones (Ellis, 1993), and even the consumer-based subculture of modern bikers who ride Harley-Davidson motorcycles (Schouten & McAlexander, 1995). In most cases, you conduct ethnographic research in field settings, which makes the ethnographer a field researcher.

As with any research method, ethnography takes a systematic approach to the topics and systems for which it is used. Berg (2009) describes the process of conducting ethnographic research in detail. According to Berg, an ethnographer faces several issues with this type of research. We explore these next.

**Observing as a Participant or Nonparticipant** One decision that you will have to make early on is whether to conduct your observations using **participant observation**, in which you act as a functioning member of the group, or using **nonparticipant observation**, in which you observe as a nonmember. In addition, you will have to decide whether to conduct your observations overtly (the group members know that you are conducting research on the group) or covertly (unobtrusive observation). When done overtly, both participant and nonparticipant observation carry the possibility of subject reactivity; as we noted in Chapter 5, group members who know they are being observed may behave differently than they otherwise would, thus

threatening external validity. This problem becomes more serious with participant observation in which you interact with your participants.

You can minimize this problem by training participant-observers not to interfere with the natural process of the group being studied or by using observers who are blind to the purposes of the study. Alternatively, if you must use participant observation, you could always become a passive participant. As such, you would keep your contributions to the group to a minimum so that you would not significantly alter the natural flow of behavior. Your main role would be to observe and record what is going on. Finally, if possible, you could become a nonparticipant-observer and avoid interacting with the group altogether.

You can reduce or remove the problem of reactivity by choosing to observe covertly. Nonparticipant covert observation is essentially naturalistic observation. Because your subjects do not know they are being observed, their behavior will be natural. Becoming a covert participant entails joining the group to be observed without disclosing your status as a researcher, so again your subjects will behave naturally in your presence. Additionally, by using covert entry, you may be able to gain access to information that would not be available if you used an overt entry strategy.

When using covert entry, you still need to be concerned about your presence disrupting the normal flow of the social interactions within the group. Your presence as a member-participant may influence how the group functions.

The practice of covertly infiltrating a group carries with it ethical liabilities. Because your subjects are not aware that they are being studied, they cannot give informed consent to participate. As discussed in Chapter 7, such violations may be acceptable if your results promise to make a significant contribution to the understanding of behavior. Thus, before deciding on covert entry, you must weigh the potential benefit of your research against the potential costs to the participants. You should adopt a covert entry strategy only if you and your institutional review board agree that the potential benefits outweigh the potential costs.

**Gaining Access to a Field Setting** Your first task when conducting ethnographic research is to gain access to the group or organization that you wish to study. In some cases, this would be easy. For example, if you wanted to conduct an ethnographic study of mall shoppers during the Christmas shopping season, you would only need to situate yourself in a mall and record the behaviors and verbalizations of shoppers. Being a public place, the mall offers free and unlimited access. In other cases, settings are more difficult to access. To conduct an ethnographic study of the police subculture, for example, you probably would need to obtain the permission of the police commissioner, various high-ranking police officers, and perhaps the rank-and-file police officers. Only then would you have access to police stations, squad cars, patrols, and so on. Access to the meeting places of elite groups (e.g., country clubs) also may be difficult because such groups often establish barriers and obstacles (such as membership requirements and restrictive guest access) to limit who has access to the facilities (Berg, 2009).

**Gaining Entry Into the Group** Gaining access to the research setting often requires gaining entry into the group that you plan to study. A popular strategy is “case and approach” (Berg, 2009). In this strategy, you first “case” the group, much as a criminal

cases a bank before robbing it. That is, you try to find information about the group, such as its social structure, its hierarchy of authority (if any), and its rituals and routines. Such foreknowledge makes it easier to enter the group and function effectively once inside. Berg suggests starting your search in the local library. Here you may find valuable information in newspapers, magazines, and other information sources. You also might check the Internet; many groups maintain Web sites that provide literature about themselves. Blogs and social networking sites such as Facebook may also provide information.

To enter the group, you may have to bargain with the members to establish your role and your boundaries (Berg, 2009). You also may find that you need to get past *gatekeepers* who serve as formal or informal protectors of the group. Getting past gatekeepers may require some negotiation and mediation (Berg, 2009). If you can cast your research in a favorable light, you may facilitate your entry into the group. For example, in an ethnographic study of prisoners in a county jail, the gatekeepers are the warden and high-ranking enforcement officials. Your chances of gaining entry into the prison population will be greater if you can convince those officials of the potential benefits of your study.

Another strategy for gaining entry into a group is to use *guides* and *informants* (Berg, 2009). These are members of the group (e.g., model inmates and correction officers) who can help convince the gatekeepers that your aims are legitimate and your study is worthwhile.

Although these techniques for gaining entry into a group are effective, they raise some ethical issues. The targets of your observations may not know that they are being studied, so participants cannot give informed consent. However, recall from Chapter 7 that under certain conditions an institutional review board (IRB) may approve a study that does not include informed consent. You need to consider the ethical implications of your ethnographic research and justify to an IRB the need to suspend the requirement for informed consent.

**Becoming Invisible** Once inside the group, your presence may alter the behavior of your participants or the operation of the social system you are studying. Berg (2009) suggests several strategies for making yourself “invisible.” If you are using an overt entry strategy, you could join in the routines and rituals of your participants, or you could foster good relations with your participants. You also could choose to enter covertly, masking your role as researcher. Whichever strategy you use, there are dangers to making yourself invisible (Berg, 2009). For example, if you use covert entry, there is the danger that your identity will be discovered, which would shatter any credibility you may have had.

**Making Observations and Recording Data** The essence of ethnography is to keep a careful record of what transpires within the group being studied. The various recording techniques discussed previously can be applied to ethnography. You could, for example, make copious notes during critical interactions. If you were riding along with police officers, you could make notes of what is said and done during routine traffic stops. When such overt note taking is not possible (especially if you have decided to use covert entry into a group), you could instead keep scraps of paper or index



cards and jot down thoughts you will expand later (Berg, 2009). You also could use voice-activated audio recorders or other recording devices. Another strategy involves waiting until the end of the day when you are alone to record your observations. One drawback to this latter strategy is that you are relying on your memory for your field notes. Over the course of the day, you may forget some details or distort others.

**Analyzing Ethnographic Data** If you take a purely qualitative approach, your data do not take the form of numbers (e.g., the number of times that a police officer threatens a suspect with arrest) but rather the form of narrative field notes from which themes and ideas are to be extracted. The first step in analyzing ethnographic data is to do an initial reading of your field notes to identify any themes and hypotheses, perhaps with an eye toward identifying themes and hypotheses overlooked (Berg, 2009). You also would systematically extract any major topics, issues, or themes present in your field records (Berg, 2009). The second step in analyzing ethnographic data is to code any systematic patterns in your notes and consider doing an in-depth content analysis (as discussed later in this chapter). Of course, this analysis strategy would be strengthened by using multiple, independent coders and content analyzers.

**An Example of Ethnography: Rationalizing Smoking** Despite the well-known dangers of smoking, millions of people around the world continue to smoke or begin smoking each year. Smokers are exposed to warnings about smoking (in public service advertisements, in messages on packs of cigars and cigarettes, and via the news media) and have been socially marginalized (e.g., relegated to smoking in designated areas only, away from nonsmokers). Despite all of this, smoking remains popular. How do people who smoke reconcile the hazards and social stigma associated with smoking and their continued smoking? One way is to rationalize. For example, smokers may come up with a list of advantages to smoking (e.g., weight control or a calming effect) to justify their decision to smoke. An ethnographic study conducted by Alan DeSantis (2003) investigated how smokers rationalized their decision to continue smoking.

DeSantis (2003) conducted his ethnographic study on cigar smokers who met regularly at a popular cigar shop in Kentucky. The study was conducted over a 3-year period (1997–2000). DeSantis used participant observation as his principal method of data collection. Cigar shop patrons were aware that DeSantis was a researcher. Before he began his study, DeSantis became a “regular” at the cigar shop, which allowed him to be “both a friend and a researcher with unlimited access to the shop’s rituals, conversations, self-disclosures, arguments, parties and weekend outings” (p. 435). DeSantis was able to make firsthand observations of the cigar patrons’ professional and private lives. He even became the drummer in the cigar shop’s rock ‘n’ roll band! DeSantis, however, was very careful not to get so close to his participants that his observations and conclusions were compromised.

DeSantis (2003) made observations an average of 2 days per week (2 hours per visit) over the 3-year period of the study, using three procedures to collect data. First, he made extensive field notes in a notebook, to which the participants quickly habituated. Second, audio recordings were made of more extensive interactions and analyzed later. Third, DeSantis made extensive postencounter field notes. That is, after an observation at the shop, he would make extensive detailed notes of the observation session.

DeSantis (2003) found that the cigar store patrons developed five rationalizations for their smoking. These five rationalizations recurred throughout the period of observation. The five recurring rationalizations are as follows:

1. Things done in moderation won't hurt you.
2. Cigar smoking is actually beneficial to one's health through stress reduction.
3. Cigars are not as bad as cigarettes.
4. Research linking cigar smoking to health consequences is flawed and therefore invalid.
5. Other hazards in life are far more dangerous than cigar smoking.

DeSantis (2003) provided qualitative data to support each rationalization. These data were actual statements made by the cigar smokers. For example, in support of the moderation argument, one patron noted that "if I smoked cigars constantly, seven days a week, had one in my mouth all the time, I would worry about it. But on most Sundays, I will not smoke at all" (p. 447). In support of the health benefits rationalization a patron said, "I am kind of a hyper guy anyway. I need something to cool me down. That is probably why I smoke" (p. 449). And, in support of the flawed research rationalization another patron said, "What they tell you today is good for you, will kill you tomorrow. I have seen too many reversals over the years" (p. 454).

DeSantis's (2003) ethnographic analysis of the rationalizations of the cigar smokers is purely qualitative. Each rationalization and the evidence to support it are described in purely verbal terms. Nowhere in our brief description of the study are any numbers, percentages, or other statistics mentioned. For example, DeSantis did not count the number of statements made in support of the "flawed research" rationalization. Instead, he provided actual quotations from the cigar smokers to support the existence of this rationalization. The ethnographic analysis provided is purely *descriptive* in nature. That is, we cannot explain *why* an individual buys into a particular rationalization.

Although DeSantis (2003) made no quantitative analyses of his data, there is nothing about ethnography that precludes at least some quantification. For example, DeSantis might have reported on the average number of cigars that patrons smoked each week. The questions that you are interested in addressing should drive your decision concerning the mix of qualitative and quantitative analyses you will use in your study.

## QUESTIONS TO PONDER

---

1. Define naturalistic observation and unobtrusive observation. How are they used to study behavior?
2. What are some of the advantages and disadvantages of naturalistic observation?
3. What is ethnography and what are the issues facing a field ethnographer?
4. How are ethnographic data recorded and analyzed?

## Sociometry

**Sociometry** involves identifying and measuring interpersonal relationships within a group (Berg, 2009). Sociometry has been applied to the systematic study of friendship patterns among children (e.g., Vandell & Hembree, 1994) and peer assessments of teenagers solicited to deal drugs (Weinfurt & Bush, 1995), as well as other social networks and work relationships (Berg, 2009).

To conduct a sociometric study, you have research participants evaluate each other along some dimension. For example, if you were interested in studying friendship patterns among third-grade students, you could have the students identify those in the class who are their friends and those who are not their friends. You could obtain similar sociometric ratings in a study of relationships among adults in a workplace.

You can use sociometry as the sole research tool to map interpersonal relationships, for example, to map friendship choices among five people in a club. (You might have each of the five people rank the three individuals they like best.) Figure 8-4 shows some hypothetical data from such a study. The individuals being chosen appear along the top; those doing the choosing along the side. For example, person A chose person B as her first choice, person E as her second, and person D as her third. Based on the data in Figure 8-4, you could graphically represent the pattern of friendship choices on a **sociogram**. Figure 8-5 displays an example sociogram.

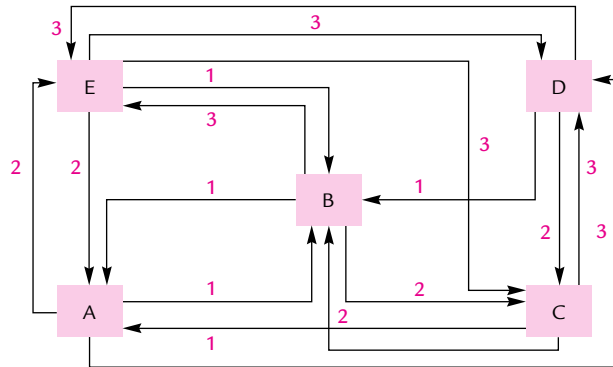
You can include sociometric ratings within a wider study as one of many measures. For example, you could use the sociometric ratings just presented in a study of whether sociometric status within a group relates to leadership roles. If there is a relationship between friendship choices and leadership roles, you would expect person B to emerge as a club leader.

**An Example of Sociometry: Peer Rejection** Evidence shows that if a child is rejected by his or her peer group it can have an effect on the child's social and emotional development as well as on his or her academic performance. A study by Christina Salmivalli, Ari Kaukiainen, and Kirsti Lagerspetz (2000) sought to determine whether peer rejection relates to the type of aggression that a child displayed and to the child's gender.

		Person chosen				
		A	B	C	D	E
Person choosing	A		1		3	2
	B	1		2		3
	C	1	2		3	
	D		1	2		3
	E	2	1		3	

**FIGURE 8-4** Example of a sociometric scoring sheet.

**FIGURE 8-5** Sociogram based on the data from Figure 8-4.



The participants, male and female ninth graders, evaluated their own and other children's aggression on a standardized measure of aggression. The participants also identified three male and three female classmates who they liked the most and three who they liked the least. This latter measure is the sociometric aspect of the study.

The results of this study showed that overall, peers who were rated as aggressive were likely to be socially rejected by both male and female children. However, lack of aggression was by no means a guarantee of social acceptance. Salmivalli et al. (2000) also found that female children were most likely to reject a peer (male or female) who used physical or verbal aggression. A different pattern emerged for male children. Male children were likely to reject a child who used verbal aggression but were less apt to reject a child who used physical aggression.

### The Case History

In some instances, your research needs may require you to study in depth a single case or just a few cases. The **case history** is a descriptive technique in which you observe and report on a single case (or a few cases). A case is the object of study, such as the development of a certain disease in a given individual.

The case history method has a long history in psychology and has many uses. For example, a case history can be used to describe the typical development of a disease or the symptoms of a new disorder. In 1861 Paul Broca reported a case history of a 51-year-old patient who died at the Bicêtre hospital in France. Broca noted that when the patient (whom he called "Tan") was admitted to the hospital at the age of 21, he had already substantially lost his capacity to speak. In fact, he would usually respond to a question with one-syllable answers, most often with the word "tan" (thus, Broca's name for him). Tan was capable of understanding what was said to him, but he could not reply. Other than the speech problem, Tan was relatively healthy. However, over the course of his hospital stay, Tan's health gradually deteriorated to the point where he was losing control over the right side of his body. After Tan died, Broca examined Tan's brain and found a syphilitic lesion in the left frontal lobe. Broca reported that "the frontal lobe of the left hemisphere was soft over a great part of its extent; the convolutions of the orbital region, although atrophied, preserved their shape; most of

the other frontal convolutions were entirely destroyed. The result of this destruction of the cerebral substance was a large cavity, capable of holding a chicken egg and filled with serous fluid" (Broca, 1861, p. 237).

Broca's case study of Tan became one of the most important findings in physiological psychology. Broca concluded that Tan's speech impairment was due to the lesion in the left frontal lobe and that the progressive damage to his brain eventually caused Tan to lose motor function on the right side of his body. Today we know that the area described by Broca (now called Broca's area) is essential to the articulation of spoken language.

Although a case history can be useful, it does not qualify as an experimental design. In fact, a case history is a special application of a demonstration (see Chapter 4). Because you do not manipulate independent variables, you cannot determine the causes of the behavior observed in your case history. You can, of course, speculate about such causes. You can even compare theories by interpreting cases from different perspectives, but you cannot state with any certainty which perspective is superior.

## Archival Research

**Archival research** is a nonexperimental strategy that involves studying existing records. These records can be historical accounts of events, census data, court records, police crime reports, published research articles, or any other archived information.

When planning archival research, you should have specific research questions in mind. You may find that the archived material contains an overwhelming amount of information. You need to be able to focus on specific aspects of the material. You can do so only if you know what you are looking for, which depends on having clearly defined and focused research hypotheses. In addition, all the factors pertaining to observational research (developing categories, coding sheets, multiple raters, etc.) apply to archival research.

An important practical matter to consider is your need to gain access to the archived material. This may not be easy. Sometimes the records you are interested in are not available to the general public. You may need to obtain special permission to gain access to the archives. In other cases, archival information may be available in libraries or on a computerized database (e.g., the Prosecutor's Management Information System, or PROMIS). Even in these cases, you may have to do some homework to find out how to access them.

Another practical matter is the completeness of the records. After gaining access to the archives, you may find that some of the information you wanted is unavailable. For example, if you were interested in studying court records, you might find that some information, such as background information on the defendant, is confidential and unavailable to you. In short, archived material may not be complete enough for your purposes. You may need to use multiple sources.

Like the case history method, archival research is purely descriptive. You may be able to identify some interesting trends or correlations based on your archival research. However, you cannot establish causal relationships.

**An Example of Archival Research: Why Don't More Women Play Chess?** Chess is one of the most intellectually challenging games that one can play. Only a very small number of individuals ever rise to the ranks of the world's top-ranked chess players, and of those who do rise to the level of chess Grandmaster only 1% are female. Chess rankings are done purely objectively, based on one's performance in chess tournaments and matches, so it is unlikely that this overwhelming disparity is due to gender discrimination (Chabris & Glickman, 2006). If it is not discrimination within the world of chess, what can explain the striking disparity? Christopher Chabris and Mark Glickman (2006) addressed this question in an archival study of the U.S. Chess Federation's (USCF) database. The USCF database served as the archival source for the data used in the study.

Chabris and Glickman (2006) examined the records of all USCF members who were active members between 1992 and 2004, a database that included over 250,000 entries. They recorded each player's birth date, gender, ZIP code, and year-end chess rating (which is an index of the player's playing strength). The first question that Chabris and Glickman addressed was whether there was a significant difference between males and females on their chess ratings. Here they found that males earned higher average ratings than females by around 500 points. Even after statistically controlling for other variables, they found a large difference between male and female average ratings (150–200 points). Although there was a large difference between mean ratings for males and females, both genders showed about the same level of variability. That is, male chess ratings were no more or less variable than female chess ratings.

Next, Chabris and Glickman (2006) tried to find out why the difference between males and females existed. First, they ruled out differential attrition rates for males and females. They found that males and females dropped out of chess at approximately the same rates. Males and females also improved their ratings at about the same rate. So, what could account for the gender difference? Chabris and Glickman found that the best explanation for the disparity was that significantly fewer females than males entered the lower levels of competitive chess. Most likely, according to Chabris and Glickman, males and females have comparable chess abilities. It is just that females have a vastly lower rate of participation, perhaps due to the lack of female role models at the highest levels of chess competition.

### Content Analysis

Use **content analysis** when you want to analyze a written or spoken record (or other meaningful matter) for the occurrence of specific categories or events (such as pauses in a speech), items (such as negative comments), or behavior (such as factual information offered during group discussion). Because it is difficult to content-analyze such materials in real time, you normally use archival sources for a content analysis. For example, if you wanted to content-analyze the answers given by two political candidates during a debate, it would be nearly impossible to do the analysis while the debate is going on. Instead, you would record the debate and use the resulting footage for your analysis. There are times, however, when you do a content analysis in real time. An example is a content analysis of court proceedings. Observers may actually sit in the courtroom and perform the content analysis.

Content analyses have been conducted on a wide range of materials such as mock juror deliberations (Horowitz, 1985), the content of television dramas (Greenberg, 1980), and the content of children's literature (Davis, 1984). In fact, the possible applications of content analysis are limited only by the imagination of the researcher (Holsti, 1969).

A relatively recent source of material for content analysis is the Internet, which offers a vast array of sources such as social networking sites (e.g., Facebook, Blogger), discussion lists, and chat rooms. Because the textual content of these sources is already stored in computer-readable files, such materials have the advantage that they can be submitted directly to specialized computer programs designed to perform content analysis. (See Krippendorff, 2004, for a thorough presentation of content analysis.)

Even though a content analysis seems rather simple to do, it can become as complex as any other research technique. You should perform content analysis within the context of a clearly developed research idea, including specific hypotheses and a sound research design. All the factors that must be considered for observational research (except that of remaining unobtrusive) apply to a content analysis. You must clearly define your response categories and develop a method for quantifying behavior. In essence, content analysis is an observational technique. However, in content analysis, your unit of analysis is some written, visual, or spoken record rather than the behavior of participants.

**Defining Characteristics of Content Analysis** Holsti (1969) points out that proper content analysis entails three defining characteristics. First, your content analysis should be objective. Each step of a content analysis should be guided by an explicit, clear set of rules or procedures. You should decide on the rules by which information will be acquired, categorized, and quantified and then adhere to those rules. You want to eliminate any subjective influence of the analyst.

Second, your content analysis should be systematic. Assign information to categories according to whatever rules that you developed and then include as much information as possible in your analysis. For example, if you are doing a content analysis of a body of literature on a particular issue (such as racial attitudes), include articles that are *not* in favor of your position as well as those that are in favor of your position. A content analysis of literature is only as good as the literature search behind it.

Third, your content analysis should have generality. That is, your findings should fit within a theoretical, empirical, or applied context. Disconnected facts generated from a content analysis are of little value (Holsti, 1969).

**Performing Content Analysis** To ensure that you acquire valid data for your content analysis, you must carefully define the response categories. According to Holsti (1969, p. 95), your categories should reflect the purposes of the research, be exhaustive, be mutually exclusive, be independent, and be derived from one classification system.

The first requirement is the most important (Holsti, 1969): clear operational definitions of terms. Your categories must be clearly defined and remain focused on the research question outlined in your hypothesis. Unclear or poorly defined categories are difficult to use. The categories should be defined with sufficient precision to allow precise categorization. However, you do not want your categories to be too narrowly

defined. You do not want relevant information to be excluded from a category simply because it does not fit an overly restrictive category definition.

Determining what your categories should be and how you should classify information within them is sometimes difficult. Reviewing related research in which a content analysis was used can help you develop and clearly define your categories. You can then add, delete, or expand categories to fit your specific research needs.

Before you begin to develop categories, read (or listen to) the materials to be analyzed. This will familiarize you with the material, help you develop categories, and help you avoid any surprises. That is, you will be less likely to encounter any information that does not fit into any category. Avoid making up categories as you go along. After developing your categories, you decide on a unit of analysis. The *recording unit* (Holsti, 1969) is the element of the material that you are going to record. The recording unit can be a word (or words), sentences, phrases, themes, and so on. Your recording unit should be relevant to your research question. Also, Holsti points out that defining a recording unit sometimes may not be enough. For example, if you were analyzing content of a jury deliberation, recording the frequency with which the word *defendant* (the recording unit) was used might not be sufficient. You might also have to note the *context unit*, or context within which the word was used (Holsti, 1969). Such a context unit gives meaning to the recording unit and may help later when you interpret the data. For example, you might record the number of times the word *defendant* was used along with the word *guilty*.

Another factor to consider when performing a content analysis is who will do the analysis. Observer bias can be a problem if the person performing the content analysis knows the hypotheses of the study or has a particular point of view. In such an instance, your results could be affected by your observer's biases. To avoid this problem, you should use a "blind" observer to do your ratings, one who does not know the purpose of your study. Also, avoid using observers who have strong feelings or characteristics that could bias the results. If you use more than one observer (and you should), you must evaluate interrater reliability.

Another important thing to remember about content analysis is that the validity of your results will depend on the materials analyzed. Make every effort to obtain relevant materials, be they books, films, or television shows. In many cases, it is not feasible to analyze all materials. For example, a content analysis of all children's books is impossible. In such cases, obtain a sample of materials that is representative of the larger population of materials. A content analysis of a biased sample (e.g., only children's books written to be nonsexist) may produce biased results.

The results from a content analysis may be interesting in and of themselves. You may discover something interesting concerning the topic under study. Such was the case with Greenberg's (1980) content analysis of prime-time television shows aired during the fall of 1977. Greenberg found that Blacks were portrayed more often as having low-status jobs and athletic physiques compared with Whites.

**Limitations of Content Analysis** Content analysis can be a useful technique to help you understand behavior. However, keep in mind that content analysis is purely descriptive. It cannot establish causal relationships among variables. Another limitation of content analysis centers on the durability of the findings. In some instances, results from a content analysis are invalidated over time. For example, Greenberg's



(1980) findings about how Blacks are portrayed on television are probably no longer valid. Currently, Blacks are more likely to be portrayed in higher-status roles (doctors, lawyers, etc.) than in the past. Of course, this prediction could be tested with an updated content analysis!

**An Example of Content Analysis: How Violent Are Video Games?** Violence seems to be all around us. News stories abound concerning school shootings, violent crimes, and interpersonal violence. One possible source of violent behavior is the media. Media critics contend that there are high levels of violence portrayed in the media (television, film, video games). Is this the case? One can answer such a question through content analysis. In fact, a content analysis by Stacy Smith, Ken Lachlan, and Ron Tamborini (2003) looked at the violent content in one media format: video games.

For this content analysis, the researchers analyzed the 20 most popular video games available for major home gaming systems (e.g., Sony PlayStation and Nintendo). The researchers classified the games as either for “mature audiences” or for “general audiences.” The first 10 minutes of each game were coded for violent content. The researchers defined violence as “any overt depiction of a credible threat of physical force or the actual use of such force intended to harm an animate being or group of beings” (Smith et al., 2003, p. 62). Smith et al. used two measures of violent content: the proportion of a video game segment that included violence and the rate of violence per minute. Smith et al. recorded the nature of the perpetrator of violence and the nature of the target.

Smith et al. (2003) found that video games intended for mature audiences contained a higher proportion of violence than those intended for the more general audience. Video games targeting mature audiences also were found to have four times as many violent acts per minute as those intended for the general audience. Overall, 68% of the video games (regardless of intended audience) had at least one act of violence. Figure 8-6 shows some of Smith et al.’s findings relating to the perpetrators and targets of violence in the video games analyzed. Both the perpetrators and the targets of violence were most likely to be White human adult males.

## QUESTIONS TO PONDER

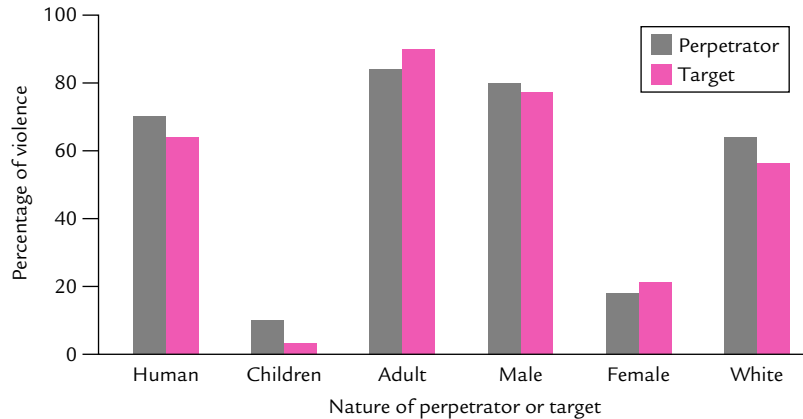
---

1. What is sociometry and when is it used?
2. How are the case history and archival research used?
3. What is content analysis, and what steps are taken when using it?

## META-ANALYSIS: A TOOL FOR COMPARING RESULTS ACROSS STUDIES

---

Imagine that you are a researcher investigating the relationship between attitudes and memory. Specifically, you have been investigating whether or not participants recall more attitude-consistent information than attitude-inconsistent information. After



**FIGURE 8-6** Results of a content analysis of violent video games.  
SOURCE: Smith, Lachlan, and Tamborini, 2003.

conducting several empirical investigations, you decide that a published literature review is needed to summarize and integrate the findings in the area. Consequently, you decide to conduct a literature review and to write a review article.

One strategy for this task is to conduct a traditional literature review. With this strategy, you read the relevant research in your area and then write an article. In your review, you may choose to summarize the major methods used to research the attitude-memory link, report the results of the major studies found, and draw conclusions about the variables that affect the relationship of interest.

In the traditional literature review, you simply summarize what you find and draw conclusions about the state of knowledge in a given area. For example, you might conclude that a certain variable is important (such as the length of a persuasive communication to which an individual is exposed) whereas others are less important (such as incidental versus intentional learning). However, the conclusions that you draw are mostly subjective, based on your critical evaluation of the literature. The possibility exists that your subjective conclusion may not accurately reflect the strength of the relationships examined in your review.

You can avoid this possibility by adding a meta-analysis to your traditional review. A **meta-analysis** is a set of statistical procedures that allow you to combine or compare results from different studies. Because you are making use of existing literature, meta-analysis is a form of archival research. When you conduct a meta-analysis, you find and analyze existing research (published and even unpublished) so that you can make statistically guided decisions about the strength of the observed effects of independent variables and the reliability of results across studies. You can also do a meta-analysis of existing meta-analyses. This technique is known as a *second-order meta-analysis* (Hunter & Schmidt, 2004). In this technique, you find as many topic-relevant meta-analyses as possible and do a meta-analysis of their results.

To conduct a meta-analysis, you must follow three steps: (1) identify relevant variables, (2) locate relevant research to review, and (3) conduct the meta-analysis proper.

### Step 1: Identifying Relevant Variables

Before you can hope to conduct a meta-analysis, you must identify the variables to be analyzed. This may sound easy enough. However, you will find that in practice it is somewhat difficult, especially in a research area in which there is a wide body of research. Generally, the rules that apply to developing testable research questions (see Chapter 2) also apply to meta-analysis. It is not enough to say, "I want to do a meta-analysis of the memory literature." Such a broad, general analysis would be extremely difficult to do. The same is true even in less extensive research areas.

Your research question must be sufficiently focused to allow for a reasonable meta-analysis. The unit of analysis in a meta-analysis should be the impact of variable *X* on variable *Y* (Rosenthal, 1984). Therefore, focus only on those variables that relate to your specific question. For example, you might choose to meta-analyze the impact of imagery on memory. Here you are limiting yourself to a small segment of the memory literature.

After you have narrowed the scope of your analysis, you must decide what variables to record (such as sex of subject, independent variables) as you review each study. Your decision will be driven by your research question. Table 8-3 provides a list of information that might be included in a meta-analysis. For each study to be included in your meta-analysis, you should record the relevant variables, the full reference citation, and the nature of the subject sample and procedure (Rosenthal, 1984).

The heart of meta-analysis is the statistical combination of results across studies. Consequently, you also must record information about the findings from the results sections of the papers you review. What information is needed depends on the meta-analytic technique that you use. To be safe, record the values of any statistics given (e.g., *t*s and *F*s) and the associated *p* values (such as .05, .01). Later these values will be used as the "scores" in your meta-analysis. You should collect data that help you evaluate your specific research questions. You do not have to record the results

**TABLE 8-3 Sample of Factors to Include When Meta-Analyzing Literature**

Full reference citation
Names and addresses of authors
Sex of experimenter
Sex of subjects used in each experiment
Characteristics of subject sample (such as how obtained, number)
Task required of subjects and other details about the dependent variable
Design of the study (including any unusual features)
Control groups and procedures included to reduce confounding
Results from statistical tests that bear directly on the issue being considered in the meta-analysis (effect sizes, values of inferential statistics, <i>p</i> values)

SOURCE: Adapted from Rosenthal, 1984.

from overall analyses of variance (ANOVAs). Focus instead on the results of statistical tests that evaluate the specific relationships among the variables of interest (Rosenthal, 1984).

### Step 2: Locating Relevant Research to Review

One of the most important steps in a meta-analysis is locating relevant research to review. In meta-analysis, you want to draw conclusions about the potency of a set of variables in a particular research area. To accomplish this end, you must thoroughly search the literature. Chapter 3 described how to perform a literature search, so the topic is not examined here.

Recall the previously discussed *file drawer phenomenon* (Chapter 3) in which studies that do not achieve statistically reliable findings fail to reach publication (Rosenthal, 1979, 1984). The problem posed by the file drawer phenomenon is potentially serious for meta-analysis because it results in a biased sample. This bias inflates the probability of making a Type I error (concluding that a variable has an effect when it does not). Studies that failed to be published because the investigated variables did not show statistically significant effects are not available to include in the meta-analysis.

There are two ways of dealing with the file drawer phenomenon. First, you can attempt to uncover those studies that never reach print. You can do this by identifying as many researchers as possible in the research area that you are covering. You then send each researcher a questionnaire, asking if any unpublished research on the issue of interest exists. You also could do a search of online journals that publish studies that produce null results such as the *Journal of Articles in Support of the Null Hypothesis* (<http://www.jasnh.com/>). Second, Rosenthal (1979, 1984) suggests estimating the extent of the impact of the file drawer phenomenon on your analysis. This is done by determining the number of studies that must be in the file drawer before serious biasing takes place (for details on how to estimate this, see Rosenthal, 1979, or Rosenthal, 1984, pp. 107–110). For example, if you determine (based on your analysis) that at least 3,000 studies must be in the file drawer before seriously biasing your results, then you can be reasonably sure that the file drawer phenomenon is not a source of bias.

### Step 3: Conducting the Meta-Analysis

When you have located relevant literature and collected your data, you are ready to apply one of the many available meta-analytic statistical techniques. Table 8-4 displays meta-analytic techniques that you can apply to the situation in which you have two studies. The first technique shows that you can compare studies. This comparison is made when you want to determine whether two studies produce significantly different effects. Essentially, doing a meta-analysis comparing studies is analogous to conducting an experiment using human or animal subjects. In the case of meta-analysis, each data point represents the results from a study rather than a subject's response. The second technique shows that you also can combine studies to determine the average effect of a variable across studies. Looking at the columns, you can evaluate

**TABLE 8-4 Meta-Analytic Techniques for Comparing and Combining Two Studies**

TECHNIQUE	COMMENTS
<b>Comparing Studies</b>	Used to determine if two studies produce significantly different results.
Significance testing	Record $p$ values from research and convert them to exact $p$ values (such as a finding reported at $p > .05$ may actually be $p = .036$ ). Used when information is not available to allow for evaluation of effect sizes.
Effect size estimation	Record values of inferential statistics (such as $F$ or $t$ , for example) along with associated degrees of freedom. Estimate effect sizes from these statistics. Preferred over significance testing.
<b>Combining Studies</b>	Used when you want to determine the potency of a variable across studies.
Significance testing	Can be used after comparing studies to arrive at an overall estimate of the probability of obtaining the two $p$ values under the null hypothesis (there is no causal relationship between the analyzed variables).
Effect size estimation	Can be used after comparing studies to evaluate the average impact across studies of an independent variable on the dependent variable.

SOURCE: Adapted from Rosenthal, 1984.

studies by comparing or combining either the  $p$  values from significance testing or effect sizes (Rosenthal, 1984).

Comparing effect sizes of two studies is more desirable than simply looking at  $p$  values (Rosenthal, 1984). This is because effect sizes provide a better indication of the degree of impact of a variable than  $p$  values do. (Remember, all the  $p$  value tells you is the likelihood of making a Type I error.) Use  $p$  values when the information needed to analyze effect sizes is not included in the studies reviewed.

### Drawbacks to Meta-Analysis

Meta-analysis can be a powerful tool to evaluate results across studies. Even though many researchers have embraced the concept of meta-analysis, others question its usefulness on several grounds. This section explores some of the drawbacks to meta-analysis and presents some of the solutions suggested to overcome those drawbacks.

**Assessing the Quality of the Research Reviewed** Chapter 3 pointed out that not all journals are created equal. The quality of the research found in a journal depends on its editorial policy. Some journals have rigorous publication standards; others may

not. This means that the quality of published research may vary considerably from journal to journal.

One problem facing the meta-analyst is how to deal with uneven quality of research. Should an article published in a nonrefereed journal be given as much weight as an article published in a refereed journal? Unfortunately, there is no simple answer to this question. Rosenthal (1984) suggests weighting articles according to quality.

There is no agreement as to the dimensions along which research should be weighted. The refereed–nonrefereed dimension is one possibility. You should exercise caution with this dimension because whether or not a journal is refereed is not a reliable indicator of the quality of published research. Research in a new area, using new methods, is sometimes rejected from refereed journals even though it is methodologically sound and of high quality. Similarly, publication in a refereed journal helps to ensure that the research is of high quality but does not guarantee it.

A second dimension along which research could be weighted is according to the soundness of methodology, regardless of journal quality. Rosenthal (1984) suggests having several experts on methodology rate each study for its quality (perhaps on a 0 to 10 scale). Quality ratings would be made twice: once after reading the method section alone and once after reading the method and results sections together (Rosenthal, 1984). The ratings would then be checked for interrater reliability and used to weight the degree of contribution of each study to the meta-analysis.

**Combining and Comparing Studies Using Different Methods** A frequent criticism of meta-analysis is that it is difficult to understand how studies with widely varying materials, measures, and methods can be compared. This is commonly referred to as the “apples-versus-oranges argument” (Glass, 1978).

Although common, this criticism of meta-analysis is not valid. Rosenthal (1984) and Glass (1978) suggest that comparing results from different studies is no different from averaging across heterogeneous subjects in an ordinary experiment. If you are willing to accept averaging across subjects, you can also accept averaging across heterogeneous studies (Glass, 1978; Rosenthal, 1984).

The core issue is not whether averaging should be done across heterogeneous studies but whether or not differing methods are related to different effect sizes. In this vein, Rosenthal (1984) points out that when a subject variable becomes a problem in research, you often “block” on that subject variable to determine how it relates to the differences that emerge. Similarly, if methodological differences appear to be related to the outcome of research, studies in a meta-analysis could be blocked on methodology (Rosenthal, 1984) to determine its effects.

**Practical Problems** The task facing a meta-analyst is a formidable one. Experiments on the same issue may use widely different methods and statistical techniques. Also, some studies may not provide the necessary information to conduct a meta-analysis. For example, Roberts (1985) was able to include only 38 studies in his meta-analysis of the attitude-memory relationship. Some studies had to be eliminated because sufficient information was not provided. Also, Roberts reports that when an article said

that  $F$  was less than 1 (as articles often do), he assigned  $F$  a value of zero. The problem of insufficient or imprecise information (along with the file drawer problem) may result in a nonrepresentative sample of research being included in your meta-analysis. Admittedly, the bias may be small, but it nevertheless may exist.

***Do the Results of Meta-Analysis Differ From Those of Traditional Reviews?*** A valid question is whether or not traditional reviews produce results that differ qualitatively from those of a meta-analysis. To answer this question, Cooper and Rosenthal (1980) directly compared the two methods. Graduate students and professors were randomly assigned to conduct either a meta-analysis or a traditional review of seven articles dealing with the impact of the sex of the subject on persistence on a task. Two of the studies showed that females were more persistent than males whereas the other five either presented no statistical data or showed no significant effect.

The results of this study showed that participants using the meta-analysis were more likely to conclude that there was an effect of sex on persistence than were participants using the traditional method. Moreover, participants doing the traditional review believed that the effect of sex on persistence was smaller than did those doing the meta-analysis. Overall, 68% of the meta-analysts were prepared to conclude that sex had an effect on persistence whereas only 27% of participants using the traditional method were so inclined. In statistical terms, the meta-analysts were more willing than the traditional reviewers to reject the null hypothesis that sex had no effect, so using meta-analysis to evaluate research may lead to a reduction in Type II decision errors, such as concluding that a variable has no effect when it does have one (Cooper & Rosenthal, 1980).

Cooper and Rosenthal (1980) also report that there were no differences between meta-analysis and traditional review groups in their abilities to evaluate the methodology of the studies reviewed. Also, there was no difference between the two groups in their recommendations about future research in the area. Most participants believed that research in the area should continue.

Finally, it is worth noting that using the statistical approach inherent in meta-analysis applies the same research strategy as doing statistical analyses of data from traditional experiments. When we obtain results of an experiment, we don't just look at ("eyeball") the data to see if any patterns or relationships exist. Instead, in most instances (there are some exceptions that we discuss in Chapter 12), we apply statistical analyses to evaluate whether relationships exist. By the same token, it can be argued that it is better to apply a statistical analysis to the results of different studies to see if significant relationships exist than to "eyeball" the studies and speculate about possible relationships.

## QUESTIONS TO PONDER

---

1. What is meta-analysis and what steps are involved in using it?
2. What are some of the issues facing you if you decide to do a meta-analysis?

## SUMMARY

---

In some situations, conducting an experiment may not be possible or desirable. In the early stages of research or when you are interested in studying naturally occurring behaviors of your subjects, a nonexperimental approach may be best.

Observational research involves observing and recording the behaviors of your subjects. This can be accomplished either in the field or in the lab and can use human participants or animal subjects. Although observational research sounds easy to conduct, as much preparation goes into an observational study as into any other study. Before making observations of behavior, you must clearly define the behaviors to be observed, develop observation techniques that do not interfere with the behaviors of your subjects, and work out a method of quantifying and recording behavior.

The frequency, duration, and intervals methods are three widely accepted ways to quantify behavior in an observational study. In the frequency method, you count the number of occurrences of a behavior within a specified period of time. In the duration method, you measure how long a behavior lasted. In the intervals method, you break your observation period into small time intervals and record whether or not a behavior occurred within each.

After you have decided how to quantify behavior, you must make some decisions about how to record your observations. Paper-and-pencil data-recording sheets provide a simple and, in most cases, adequate means of recording behavior. In some situations (such as when the behavior being observed is fast paced), you should consider using electronic recorders rather than a paper-and-pencil method. Using a recorder allows observers to keep their eyes on subjects while making notes about behavior.

In addition to developing a method for quantifying behavior, you must decide on how and when to make observations. Sometimes it is not possible to watch and record behaviors simultaneously because behavior may occur quickly and be highly complex. In such situations, you could use time sampling or individual sampling or automate your observations by using a video recorder.

In observational research, you should use multiple observers. When multiple observers are used, you must evaluate the degree of interrater reliability. This can be done using either percent agreement, Cohen's Kappa, intraclass correlation, or Pearson  $r$ . A Cohen's Kappa of .70 or greater or a statistically significant Pearson  $r$  of around .90 or greater suggests an acceptable level of interrater reliability.

Nonexperimental techniques include naturalistic observation, ethnography, case study, archival research, and content analysis. In naturalistic observation, you make careful, unobtrusive observations of subjects in their natural environment so that you do not alter their natural behavior. In cases in which you cannot remain unobtrusive, there are steps you can take to habituate your participants to your presence.

Ethnography involves getting immersed in a behavioral or social system to be studied. The technique is best used to study and describe the operation of groups and the social interactions that take place within those groups. An ethnographic study can be run as a participant observation, in which the researcher actually becomes a member of the group, or as nonparticipant observation, in which the researcher is a nonparticipating observer.



Sociometry involves identifying and measuring interpersonal relationships within a group. Research participants evaluate each other along some socially relevant dimension (e.g., friendship), and patterns of those ratings are analyzed to characterize the social structure of the group. The results of a sociometric analysis can be plotted on a sociogram, which graphically represents the social connections between participants. Sociometry can be used as a stand-alone research technique or as a measure within a wider study.

When using the case history approach, you analyze an interesting case that illustrates some empirical or theoretical point. Alternatively, you may compare and contrast two or more cases in order to illustrate such points. Archival research makes use of existing records. You examine those records and extract data to answer specific research questions.

Content analysis involves analyzing a written or spoken record (or other content) for the occurrence of specific categories of events or behaviors. As with any observational technique, you must develop behavior categories. During content analysis, you note and analyze recording and context units.

Meta-analysis is a family of statistical techniques that can help you evaluate results from a number of studies in a given research area. In contrast to a traditional literature review (in which subjective evaluations rule), meta-analysis involves statistically combining the results from a number of studies. Meta-analytic techniques tend to be more objective than traditional literature review techniques.

The three steps involved in conducting a meta-analysis are (1) identifying relevant variables to study, (2) locating relevant research to review, and (3) actually doing the meta-analysis (comparing or combining results across studies). Although meta-analysis has advantages over traditional literature reviews, there are some drawbacks. First, it is sometimes difficult to evaluate the quality of the research reviewed. Second, studies in a research area may use vastly different methods, making comparison of results suspect. Third, the information in published articles may be incomplete, eliminating potentially important studies from the analysis.

## KEY TERMS

---

behavioral categories

interrater reliability

Cohen's Kappa

intraclass correlation coefficient ( $r_I$ )

quantitative data

qualitative data

naturalistic observation

ethnography

participant observation

nonparticipant observation

sociometry

sociogram

case history

archival research

content analysis

meta-analysis