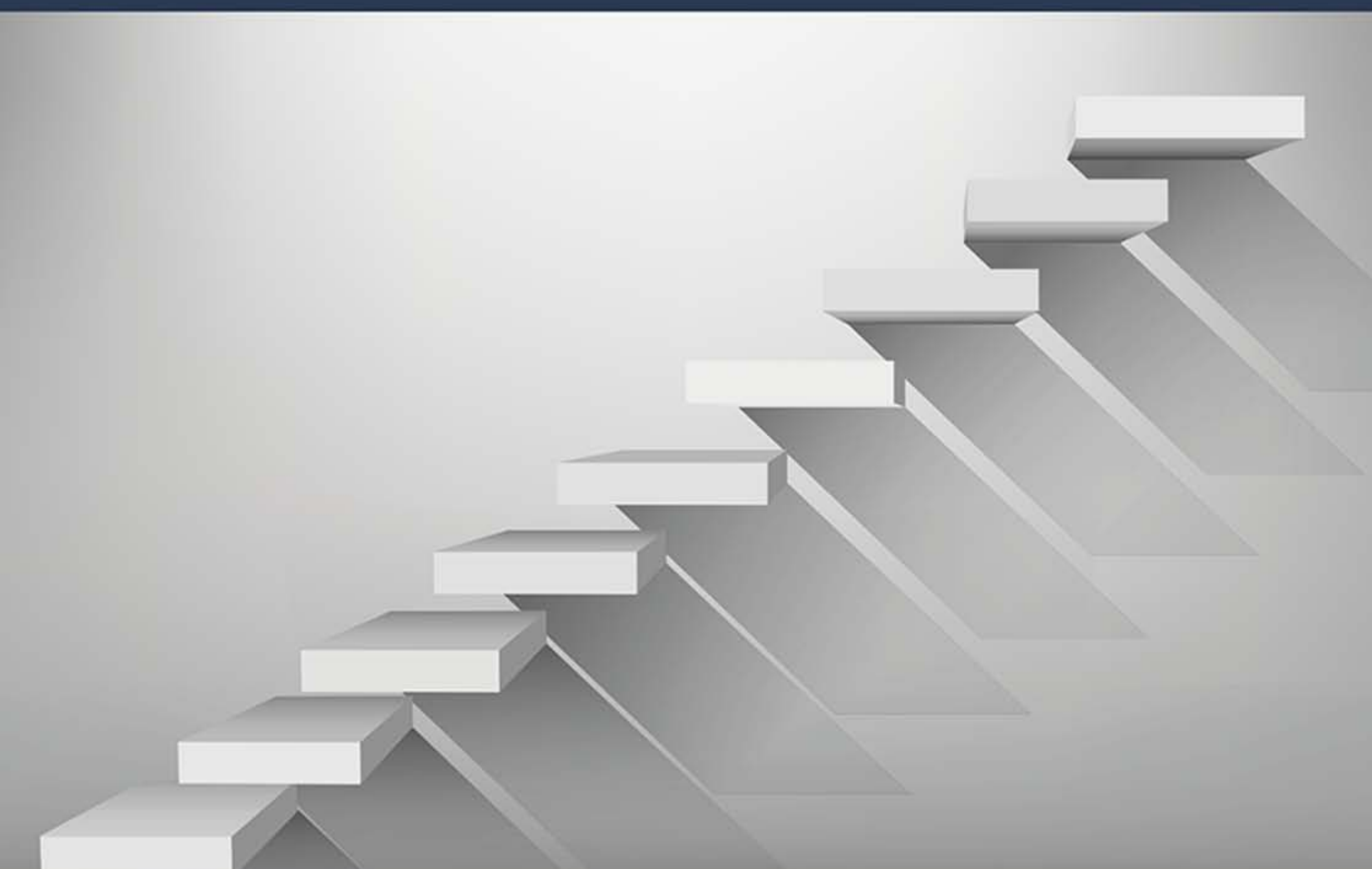
sixth edition

Public Policy Analysis

An Integrated Approach



William N. Dunn

"Unique among textbooks, Dunn puts problematization and problem structuring center stage as the major driver in each field of the policy-analytic domain. As usual, Dunn skillfully blends (social) scientific insight, epistemological sophistication, methodological finesse and practical savvy."

Robert Hoppe, *University of Twente, The Netherlands*

"This new edition continues the tradition of excellence established by previous editions and updates the text, references and examples to ensure its continuing relevance to the contemporary world of policy analysis in governments, the NGO and private sectors."

Michael Howlett, *Simon-Fraser University, Canada*

"William Dunn effectively illustrates methodology and methods of policy analysis with real-world case materials and examples at the local, national and international levels. It remains the best guide for both undergraduate and graduate courses out there."

Keun Namkoong, *Former President of Seoul Tech and of Korean Association for Public Administration, South Korea*

"Very impressive book and will continue to use this as a must read book for my students aspiring to become respectable policy analysts."

Sofian Effendi, *Gadjah Mada University, Indonesia*

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Public Policy Analysis

Public Policy Analysis, the most widely cited book on the subject, provides students with a comprehensive methodology of policy analysis. It starts from the premise that policy analysis is an applied social science discipline designed for solving practical problems facing public and nonprofit organizations. This thoroughly revised sixth edition contains a number of important updates:

- Each chapter includes an all-new "big ideas" case study in policy analysis to stimulate student interest in timely and important problems.
- The dedicated chapter on evidence-based policy and the role of field experiments has been thoroughly rewritten and expanded.
- New sections on important developments in the field have been added, including using scientific evidence in public policymaking, systematic reviews, meta-analyses, and "big data."
- Data sets to apply analytical techniques are included online as IBM SPSS 23.0 files and are convertible to Excel, Stata, and R statistical software programs to suit a variety of course needs and teaching styles.
- All-new PowerPoint slides are included to make instructor preparation easier than ever before.

Designed to prepare students from a variety of academic backgrounds to conduct policy analysis on their own, without requiring a background in microeconomics, *Public Policy Analysis, Sixth Edition* helps students develop the practical skills needed to communicate findings through memos, position papers, and other forms of structured analytical writing. The text engages students by challenging them to critically analyze the arguments of policy practitioners as well as political scientists, economists, and political philosophers.

William N. Dunn is Professor in the Graduate School of Public and International Affairs at the University of Pittsburgh, USA.

Public Policy Analysis

An Integrated Approach

Sixth Edition

WILLIAM N. DUNN

Routledge
Taylor & Francis Group
NEW YORK AND LONDON

Sixth edition published 2018

by Routledge
711 Third Avenue, New York, NY 10017

and by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

Routledge is an imprint of the Taylor & Francis Group, an informa business

First edition published by Pearson Education, Inc. 1994
Fifth edition published by Routledge 2016

Library of Congress Cataloging-in-Publication Data
Names: Dunn, William N., author.
Title: Public policy analysis : an integrated approach / by William N. Dunn.
Description: Sixth Edition. | New York : Routledge, 2017. | Revised edition of the author's Public policy analysis, c2012. | Includes bibliographical references and index.
Identifiers: LCCN 2017008665 | ISBN 9781138743830 (hardback : alk. paper) | ISBN 9781138743847 (pbk. : alk. paper) | ISBN 9781315181226 (ebook)
Subjects: LCSH: Policy sciences. | Political planning—Evaluation.
Classification: LCC H61 .D882 2017 | DDC 320.6—dc23
LC record available at https://lccn.loc.gov/2017008665

ISBN: 978-1-138-74383-0 (hbk)
ISBN: 978-1-138-74384-7 (pbk)
ISBN: 978-1-315-18122-6 (ebk)

Typeset in Minion Pro, Bell Gothic
by Apex CoVantage, LLC

Visit the eResources: www.routledge.com/9781138743847

In memory of Donald T. Campbell

BRIEF CONTENTS

BOX CONTENTS

USEFUL CONCEPTS, GUIDES, AND CHECKLISTS

CASE STUDY CONTENTS

LESSONS FROM PRACTICE

VISUAL DISPLAY CONTENTS

FIGURES FOR ORGANIZING THINKING

TABLES FOR ORGANIZING THINKING

DETAILED CONTENTS

Part III Methods of Policy Communication 347

CHAPTER 8

Developing Policy Arguments 348

CHAPTER 9

Communicating Policy Analysis 391

PREFACE

My aim in writing the several editions of this book has been to produce a critical synthesis of the field, while at the same time offering students, instructors, and practitioners a body of knowledge and skills that is applicable to real-world problems. This is not a text in the narrow and conventional sense of the term.

Ever since the publication of the first edition of Public Policy Analysis more than thirty-five years ago, I have become more and more convinced that the methodology of policy analysis rests or should rest on epistemological foundations that differ from those of the disciplines of which policy analysis is composed. For this reason, I continue to define policy analysis as an applied social science discipline that employs multiple methods of inquiry to solve practical problems.

What this means is that the methodology of policy analysis cannot be reduced to the theories and analytical routines of microeconomics, because solutions for practical problems demand much more than the analysis of rational choice, expected utility, and opportunity costs. By the same token, the methodology of policy analysis cannot be reduced to the study of politics, because solutions for practical problems require more than the analysis of power, rule, and authority or who gets what, when, and how. Much more is involved. Finally, because a principal aim of policy analysis is to improve policies, the methodology of policy analysis cannot be reduced to an academic spectator sport in which knowledge is prized for its own sake.

In this 6th edition, I have tried to employ a simplified style of writing, choose cases based on real-world analytical practices, and create visual displays that make complex ideas understandable. Case materials now include issues in foreign policy and international security as well as domestic issues of environmental justice, urban economics, transportation, health, and traffic safety. Advanced graphics for mapping and evaluating policy arguments are included in order to cultivate critical thinking skills in areas ranging from expert forecasting and statistical analysis to theories of environmental justice. The book provides students with marketable skills in communicating policy analysis by writing policy memos, position papers, and other products of structured analytical writing. Finally, new review questions and demonstration exercises emphasize active rather than passive learning.

Special instructional devices and learning strategies are again employed throughout the book:

- *Advance organizers*. The book uses advance organizers, especially visual displays, to introduce students to the rationale and application of methods. The advance organizer for the book as a whole is the functional model of policy analysis presented in the first chapter.

- *Learning objectives*. At the beginning of each chapter is a statement of the learning objectives that students should be able to attain by reading the chapter and completing its study questions and demonstration exercises. These objectives are stated in terms of the acquisition of knowledge and skills to do or perform something—that is, behaviorally anchored learning objectives that involve recognizing, defining, understanding, explaining, predicting, evaluating, and applying. By stating learning objectives in this way, the emphasis is on active rather than passive learning, on application rather than memorization.
- *Review questions*. Knowledge and skills must be reinforced. For this reason, review questions are provided at the end of each chapter. The review questions address higher-order knowledge and skills (e.g., explaining or applying) as well as lower-order knowledge and skills (e.g., calculating or estimating). Review questions may be used by students for self-study and by instructors who are developing written assignments, examinations, and tests.
- *Demonstration exercises*. Knowledge and skills are not acquired or retained without frequent opportunities for application to real-world problems. For this reason, each chapter contains opportunities to demonstrate the application of knowledge and skills to practical problems. The attempt is to draw students away from "blackboard policy analysis" into the real world of messy problems.
- *Cases*. Cases in policy analysis are the focus of the demonstration exercises. Cases span a number of issue areas including foreign policy and counter-terrorism, transportation policy, occupational health and safety, and urban policy. Some cases are primarily conceptual; most are methodological in nature.
- *Bibliographies*. In addition to literature cited in footnotes, each chapter is accompanied by a bibliography keyed to the subject matter of that chapter. I have attempted to include literature that is representative of many of the most important recent and classical developments in public policy analysis.
- *Guidelines for written and oral communication*. Students who master methods almost always face difficulties when they are faced with translating and communicating the results of analysis. For this reason, there is an emphasis on writing issue papers, policy memoranda, and position papers, and planning oral briefings. Appendices present step-by-step guidelines and checklists.
- *Argument maps*. Analysis is about making and understanding policy arguments. This edition uses argument maps created with Rationale, an innovative argument mapping program available at www.reasoninglab.com. Elsewhere I have expressed my gratitude to Professor Tim van Gelder, one of the originators of the program, for his help.
- *Website*. A special website (www.routledge.com/9781138743847) supports users of this book by providing slides keyed to each chapter and data sets related to cases covered in the chapters. Many of the slides can serve as teaching notes.
- *Argument Mapping Software*. Users of this book may wish to purchase argument mapping software (*Rationale 2*) by going to www.reasoninglab.com. Educational discounts are available.

ACKNOWLEDGEMENTS

I would not have begun this ambitious multidisciplinary project without the encouragement of the late Paul F. Lazarsfeld, who many years ago challenged me to investigate what he called the "policy sciences movement." Lazarsfeld, one of a handful of premier applied social scientists of the twentieth century, was skeptical about the breadth of the enterprise, as it had been sketched by Harold Lasswell, Daniel Lerner, and others. Its aims seemed to him unwisely all-encompassing and grand, a criticism also held by Charles Lindblom and a number of other policy specialists. Lazarsfeld, it should be said, did not self-identify as a sociologist, but as an applied social scientist with multidisciplinary commitments. For this reason, he was University Professor of Social Science (not Sociology) at the University of Pittsburgh.

Some ten years later, we made an ultimately unsuccessful attempt to fill Lazarsfeld's vacant chair with another premier applied social scientist, Donald T. Campbell, who had virtually revolutionized the methodology of the applied social sciences in the twentieth century. Campbell, like Lazarsfeld, did not self-identify primarily with his original discipline, which was social psychology, but viewed himself as a multidisciplinary applied social scientist specializing in program and policy evaluation, as well as the philosophy and sociology of science. Had he joined us, Campbell would have been a University Professor of Social Science (not Psychology) at the University of Pittsburgh.

Campbell's mentorship at a critical stage in my professional life some thirty years ago had a profound effect on the way I think about the strengths and limitations of policy analysis, program evaluation, and the applied social sciences. His imprint can be seen throughout this book.

At about the same time, I was asked to join a team of faculty who were developing curricular materials on policy analysis for practitioners in local governments. Under the leadership of Blue Wooldridge of the National Technical Information Service, now with Virginia Commonwealth University, the group had wisely contracted specialists in learning theory and curriculum development, including Doris Gow and Jyotsna Vasudev, who were brought into the project. I learned from them the important pedagogical lesson that abstract subjects such as policy analysis can be more effectively taught by focusing on behaviorally defined learning objectives and what universities now teach under the rubric of distance learning. I am grateful to them for making me see that much of the literature we assign in courses is not easily or successfully tied to learning outcomes, which means that we often cannot say why we want students to read the materials we assign. This was an embarrassing revelation.

I have been fortunate to meet and work with colleagues who changed my mind about many things, including the important role that the philosophy and sociology of science play in the applied social sciences. These colleagues include Ian I. Mitroff, Burkart Holzner, and my former student and now professor, Bahman Fozouni, who introduced me to yet other colleagues in the Center for History and Philosophy of Science at the

University of Pittsburgh. I learned much about the pragmatic importance of policy analysis and the applied social sciences from Gerald Zaltman, Robert F. Rich, Thomas D. Cook, and the late Carol H. Weiss, all of whom were and are committed to investigating and deliberately changing the conditions under which the social sciences may be used to solve practical problems. B. Guy Peters also taught me much about the need to link policy analysis with the field of governance and, particularly, policy design.

Faculty and students in the Graduate School of Public and International Affairs (GSPIA), University of Pittsburgh, have helped me improve my thinking, writing, and teaching. They include Alex Weilenmann, Louise Comfort, Dipak Gupta, Steve Coulthart, Tom Pavlak, John Mendeloff, Ilia Murtazashvili, Jennifer Murtazashvili, Hector Correa, Michael Sabath, Soumana Sako, Sam Overman, Tony Cahill, Mary Jo Dukes, Sofian Effendi, Kevin Kearns, S. Laurel Weldon, Dave Miller, Ralph Bangs, Jan Jernigan, Neda Milevska, Phil Murphy, Keun Namkoong, Andrea Hegedus, and Phil Williams.

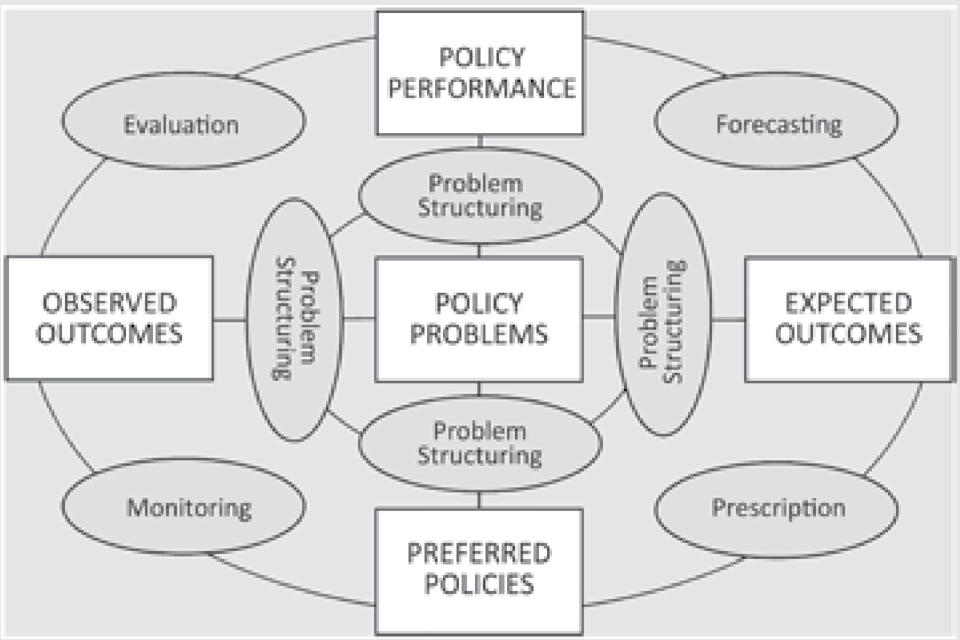
In the past thirty-five years, this book has been used in training, certificate, and degree programs in universities, think tanks, and governments in this country and abroad. Translations into Arabic, Chinese, Indonesian, Korean, Macedonian, Romanian, Russian, Ukrainian, and Spanish are completed or under way. The book has been used in training programs and projects in countries of the European Union, Southeastern Europe, the Middle East, North Africa, and Latin America. The revisions incorporated in this edition reflect some of what I have learned from the participants and organizers of these programs.

I am also grateful to students and faculty at the Graduate Center for Public Policy and Management in Skopje, Macedonia, an innovative but now inactive institution with some of the best students I have ever taught. I also want to thank reviewers of this and previous editions. In addition to anonymous reviewers, they include David Nice of Washington State University, David Houston of the University of Tennessee at Knoxville, Louise Comfort of the University of Pittsburgh. Sheila Kelly, Lien Rung-Kao, Sujatha Raman, Eric Sevigny, Kate Freed, Bojana Aceva-Andonova, Erin McGrath, Alla Golovina Khadka, and Jessica Reyes assisted in preparing materials for this and previous editions. I am also grateful for the editorial assistance of Integra Software Services Pvt Ltd, in Pondicherry, India, and the editors at Pearson. For this sixth edition I owe special thanks to Misha Kydd, Yvette M. Chin, Sophie Watson, and the publisher at Routledge, Laura D.H. Stearns.

William N. Dunn
Graduate School of Public and International Affairs,
University of Pittsburgh

PART
I

Methodology of Policy Analysis



CHAPTER 1

The Process of Policy Analysis

LEARNING OBJECTIVES

By studying this chapter, you should be able to:

- Define and illustrate phases of policy analysis
- Distinguish four strategies of policy analysis
- Contrast reconstructed logic and logic-in-use
- Distinguish prospective and retrospective policy analysis
- Distinguish problem structuring and problem solving
- Describe the structure of a policy argument and its elements
- Understand the role of argument mapping in critical thinking
- Interpret scorecards, spreadsheets, influence diagrams, decision trees, and argument maps

INTRODUCTION

Policy analysis is a process of multidisciplinary inquiry aiming at the creation, critical assessment, and communication of policy-relevant knowledge. As a problem-solving discipline, it draws on social science methods, theories, and substantive findings to solve practical problems.[1]

METHODOLOGY OF POLICY INQUIRY

The *methodology* of policy inquiry refers to the critical investigation of potential solutions to practical problems. Abraham Kaplan, one of the founders of the policy sciences, observed that the aim of methodology is to help understand and question, not only the products of policy inquiry, but the processes employed to create these products. The methodology of policy inquiry contributes to the reflective understanding of theories, methods, and practices of specialized fields such as benefit–cost analysis in economics, implementation analysis in political science, and program budgeting in public administration. These and other specialized fields hold an important place in the process of policy inquiry because they help provide its multidisciplinary foundations.

The methodology of policy inquiry is not the same as methods such as regression analysis, ethnographic interviewing, or benefit-cost analysis, because methodology is concerned with the philosophical assumptions that justify the use of these methods. Nor is the methodology of policy inquiry equivalent to one or another philosophy of science, for example, logical positivism, hermeneutics, or pragmatism.[2] On the contrary, the methodology of policy inquiry is productively eclectic; its practitioners are free to choose among a range of scientific methods, qualitative as well as quantitative, and philosophies of science, so long as these yield reliable knowledge.[3] In this context, policy analysis includes art and craft, which

[1]For a sample of alternative definitions see Harold D. Lasswell, *A Pre-view of Policy Sciences* (New York: American Elsevier Publishing, 1971); Yehezkel Dror, *Ventures in Policy Sciences: Concepts and Applications* (New York: American Elsevier Publishing, 1971); Edward S. Quade, *Analysis for Public Decisions*, 3rd rev. ed., ed. Grace M. Carter (New York: North Holland Publishing, 1989); David L. Weimer and Aidan R. Vining, *Policy Analysis: Concepts and Practice*, 2nd ed. (Englewood Cliffs, NJ: Prentice Hall, Inc., 1992).

[2]Logical positivism (or logical empiricism) was abandoned by most philosophers of science more than 50 years ago, although its epistemological pillars—the correspondence theory of truth, the empirical criterion of meaning, and value-free science—are still venerated by many social scientists. For alternatives to logical positivism in economics and political science see Daniel Bromley, *Sufficient Reason: Volitional Pragmatism and the Meaning of Economic Institutions* (Princeton, NJ: Princeton University Press, 2006); Henry E. Brady and David Collier, eds. *Rethinking Social Inquiry: Diverse Tools, Shared Standards* (Lanham, MD: Rowman Littlefield, 2004); Deirdre N. McCloskey, *The Rhetoric of Economics*, 2nd ed. (Madison: University of Wisconsin Press, 1998); Thomas Ziliak and Deirdre N. McCloskey, *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives* (Ann Arbor: University of Michigan Press, 2008).

[3]Larry Laudan has argued persuasively that the demarcation between science and nonscience, including science, art, and craft, is a pseudo-problem. It is not necessary to ask whether knowledge is “scientific,” only whether it is reliable. “The Demise of the Demarcation Problem,” in R. S. Cohen and L. Laudan, *Physics, Philosophy and Psychoanalysis: Essays in Honor of Adolf Grünbaum.* Boston Studies in the Philosophy of Science, vol. 76 (Dordrecht: D. Reidel, 1983), pp. 111–127. Aaron Wildavsky and others seem to juxtapose art and craft to characterize policy analysis. See Aaron Wildavsky, *Speaking Truth to Power: The Art and Craft of Policy Analysis* (Boston, MA: Little Brown, 1979); and Iris Geva-May and Aaron Wildavsky, *An Operational Approach to Policy*

may be regarded as scientific to the extent that they succeed in producing knowledge that is *reliable* because it can be trusted. Ordinary commonsense knowing and well-winnowed practical wisdom, both products of evolutionary learning across generations of problem solvers, often produce conclusions that are at least as reliable, and sometimes more so, than those produced by specialized social science methods.[4]

The rationale for policy analysis is pragmatic. For this reason, policy analysis is unmistakably different from social science disciplines that prize knowledge for its own sake. The policy-relevance of these disciplines depends, not only on their status as sciences, but on the extent to which they are successful in illuminating and alleviating practical problems. Practical problems, however, do not arrive in separate disciplinary packages addressed, as it were, to social science departments. In today's world, multidisciplinary policy analysis seems to provide the best fit with the manifold complexity of public policymaking.

MULTIDISCIPLINARY POLICY ANALYSIS

Policy analysis is partly *descriptive*. It relies on traditional social science disciplines to describe and explain the causes and consequences of policies. But it is also *normative,* a term that refers to value judgments about what ought to be, in contrast to descriptive statements about what is.[5] To investigate problems of efficiency and fairness, policy analysis draws on normative economics and decision analysis, as well as ethics and other branches of social and political philosophy, all of which are about what *ought* to be. This normative commitment stems from the fact that analyzing policies demands that we choose among desired consequences (ends) and preferred courses of action (means). The choice of ends and means requires continuing tradeoffs among competing values of efficiency, equity, security, liberty, democracy, and enlightenment.[6] The importance of normative reasoning in policy analysis was well stated by a former undersecretary in the Department of Housing

Analysis: The Craft, Prescriptions for Better Analysis (Boston, MA: Kluwer, 1997). The term *policy sciences* is Harold Lasswell's. See the short methodological history of the policy sciences in Ronald Brunner, "The Policy Movement as a Policy Problem," in *Advances in Policy Studies since 1950,* vol. 10, *Policy Studies Review Annual,* ed. W. N. Dunn and R. M. Kelly (New Brunswick, NJ: Transaction Books, 1992), pp. 155–197, and contributions to Michael Moran, Martin Rein, and Robert E. Goodin, eds., *The Oxford Handbook of Public Policy* (Oxford: Oxford University Press, 2006).

[4]On the contrasts between scientific and professional knowledge, on one hand, and ordinary commonsense knowing, on the other, see Charles E. Lindblom and David K. Cohen, *Usable Knowledge: Social Science and Social Problem Solving* (New Haven, CT: Yale University Press, 1979). On the overall soundness of evolved practical knowledge—but the periodic need for supplemental scientific testing—see Donald T. Campbell, "Evolutionary Epistemology," in *Methodology and Epistemology for Social Science: Selected Papers*, ed. E. S. Overman (Chicago, IL: University of Chicago Press, 1989).

[5]A classic statement of the difference between positive and normative knowledge in economics is Milton Friedman, *Essays in Positive Economics* (Chicago, IL: University of Chicago Press, 1953). This same positive-normative distinction is present throughout the social sciences. This same positive-normative distinction is the basis for Dennis C. Muller's *Public Choice III* (Cambridge: Cambridge University Press, 2012), a widely respected treatise on positive and normative economic theory and research.

[6]See Deborah Stone, Policy Paradox: *The Art of Political Decision Making*, rev. ed. (New York: W. W. Norton, 2001), and Harold D. Lasswell, *A Pre-view of Policy Sciences* (New York: Elsevier, 1971).

and Urban Development: "Our problem is not to do what is right. Our problem is to know what is right."[7]

Policy-Relevant Knowledge

Policy analysis is designed to provide *policy-relevant knowledge* about five types of questions:

- *Policy problems.* What is the problem for which a potential solution is sought? Is global warming a man-made consequence of vehicle emissions, or a consequence of periodic fluctuations in the temperature of the atmosphere? What alternatives are available to mitigate global warming? What are the potential outcomes of these alternatives and what is their value or utility?
- *Expected policy outcomes.* What are the expected outcomes of policies designed to reduce future harmful emissions? Because periodic natural fluctuations are difficult if not impossible to control, what is the likelihood that emissions can be reduced by raising the price of gasoline and diesel fuel or requiring that aircraft use biofuels?
- *Preferred policies.* Which policies should be chosen, considering not only their expected outcomes in reducing harmful emissions, but the value of reduced emissions in terms of monetary costs and benefits? Should environmental justice be valued along with economic efficiency?
- *Observed policy outcomes.* What policy outcomes are observed, as distinguished from the outcomes expected before the adoption of a preferred policy? Did a preferred policy actually result in reduced emissions, or did decreases in world petroleum production and consequent increases in gasoline prices and reduced driving also reduce emissions?
- *Policy performance.* To what extent has policy performance been achieved, as defined by valued policy outcomes signaling the reduction of global warming through emissions controls? To what extent has the policy achieved other measures of policy performance, for example the reduction of costs of carbon emissions and global warming to future generations?

Answers to these questions yield these five types of policy-relevant knowledge, which are shown as rectangles in Figure 1.1.[8]

Policy problems are representations of problem situations, which are diffuse sets of worries, inchoate signs of stress, or surprises for which there is no apparent solution. Knowledge of what problem to solve requires knowledge about the antecedent conditions of a problem situation (e.g., school dropouts as an antecedent of unemployment), as well as knowledge about values (e.g., safe schools or a living wage) whose achievement may lead to the definition of the problem and its potential solutions. Knowledge about policy

[7]Robert C. Wood, quoting President Lyndon B. Johnson in his "Foreword" to *The Study of Policy Formation*, ed. Raymond A. Bauer and Kenneth J. Gergen (New York: Free Press, 1968), p. v.

[8]A similar framework, designed for the discipline of sociology, was suggested by Walter Wallace, *The Logic of Science in Sociology* (Chicago, IL: Aldine Books, 1971). Wallace's framework addresses research methodology in sociology, while Figure 1.1 addresses the multidisciplinary research methodology of policy analysis.

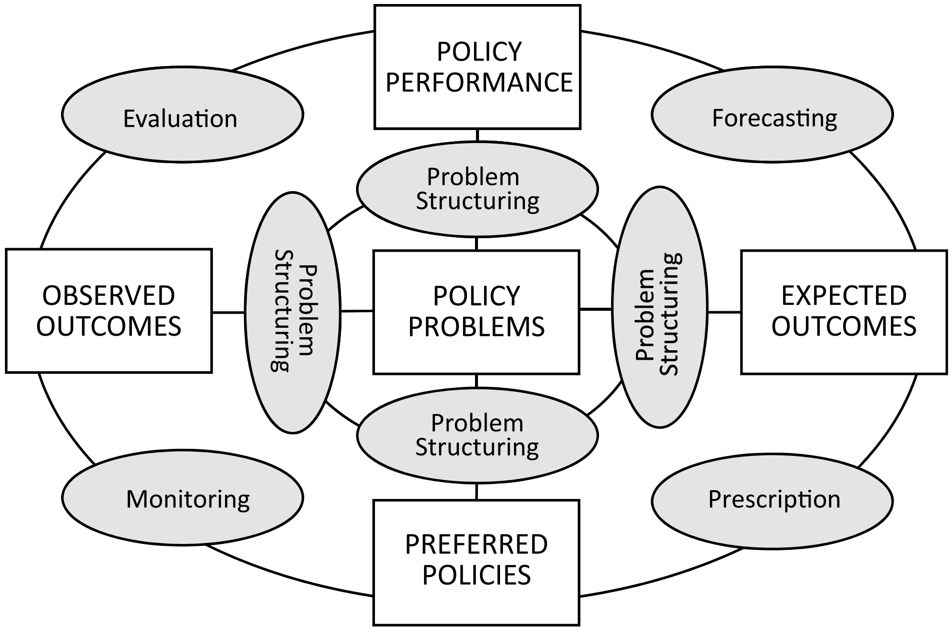


FIGURE 1.1
Multidisciplinary Policy Analysis

problems also includes at least two potential solutions to the problem and, if available, the probabilities that each alternative is likely to achieve a solution. Knowledge about policy problems plays a critical role in policy analysis, because the way a problem is defined shapes the identification of available solutions. Inadequate or faulty knowledge may result in serious or even fatal errors: defining the wrong problem.[9]

Expected policy outcomes are likely consequences of adopting one or more policy alternatives designed to solve a problem. Knowledge about the circumstances that gave rise to a problem is important for producing knowledge about expected policy outcomes. Such knowledge is often insufficient, however, because the past does not repeat itself, and the values that shape behavior may change in future. For this reason, knowledge about expected policy outcomes is not "given" by the existing situation. To produce such knowledge may require creativity, insight, and the use of tacit knowledge.[10]

[9]Defining the wrong problem is a Type III error, as contrasted with Type I and Type II errors committed when the level of statistical significance (*alpha*) is set too high or too low in testing the null hypothesis. Early statements of this contrast are Howard Raiffa, *Decision Analysis* (Reading, MA: Addison-Wesley, 1968), p. 264, and Ian I. Mitroff and Thomas R. Featheringham, "On Systematic Problem Solving and the Error of the Third Kind," *Behavioral Sciences* 19, 6 (1974): 383–393.

[10]Books that address creativity, insight, and tacit knowledge are Yehezkel Dror, *Ventures in Policy Sciences: Concepts and Applications* (New York: American Elsevier Publishing, 1971); Sir Geoffrey Vickers, *The Art of Judgment: A Study of Policy Making* (New York: Basic Books, 1965); and C. West Churchman, *The Design of*

A *preferred policy* is a potential solution to a problem. To select a preferred policy, it is necessary to have knowledge about expected policy outcomes as well as knowledge about the value or utility of the expected outcomes. Another way to say this is that factual as well as value premises are required for a policy prescription. The fact that one policy is more effective or efficient than another does not alone justify the choice of a preferred policy. Factual premises must be joined with value premises involving equality, efficiency, security, democracy, enlightenment, or some other value.

An *observed policy outcome* is a present or past consequence of implementing a preferred policy. It is sometimes unclear whether an outcome is actually an effect of a policy. Some effects are not *policy* outcomes, because many outcomes are the result of other, extra-policy factors. It is important to recognize that the consequences of action cannot be fully stated or known in advance, which means that many consequences are neither anticipated nor intended. Fortunately, knowledge about observed policy outcomes can be produced after policies have been implemented.

Policy performance is the degree to which an observed policy outcome contributes to the solution of a problem. In practice, policy performance is never perfect. Problems are rarely "solved"; most often problems are resolved, reformulated, and even "unsolved."[11] To know whether a problem has been solved, requires knowledge about observed policy outcomes, as well as knowledge about the extent to which these outcomes contribute to the valued opportunities for improvement that gave rise to a problem.

Knowledge Transformations

The five types of policy-relevant knowledge are interdependent. The solid lines connecting each pair of components in Figure 1.1 represent *knowledge transformations,* where one type of knowledge is changed into another, so that the creation of knowledge at any point depends on knowledge produced in an adjacent (and most often previous) phase. Knowledge about policy performance, for example, depends on the transformation of prior knowledge about observed policy outcomes. The reason for this dependence is that any assessment of how well a policy achieves its objectives assumes that we already have reliable knowledge about the outcomes of that policy. Note, however, that types of knowledge are connected with solid lines rather than arrows to show that knowledge can be transformed backward and forward in an iterative fashion. The process of transformation is rarely linear.

Knowledge about policy problems is a special case. Knowledge about policy problems contains other types of knowledge. Some types of knowledge may be included—for example, knowledge about preferred policies—and others excluded. What is included or excluded in the formulation of a problem affects which policies are eventually prescribed, which values are chosen as criteria of policy performance, and which expected outcomes

Inquiring Systems; Basic Concepts of Systems and Organization (New York: Basic Books, 1971). The concept of tacit knowledge is attributed to Michael Polanyi, *Personal Knowledge* (Chicago, IL: University of Chicago Press, 1958).

[11]Russell L. Ackoff, "Beyond Problem Solving," *General Systems* 19 (1974): 237–239.

warrant or do not warrant attention. At the risk of being repetitious, it is worth stressing that a fatal error of policy analysis is a Type III error—defining the wrong problem.[12]

Policy-Analytic Methods

The five types of policy-relevant knowledge are produced and transformed by using policy-analytic methods, which are the vehicles driving the production and transformation of knowledge. Methods involve judgments of different kinds[13]: judgments to accept or reject an explanation, to affirm or dispute the rightness or wrongness of a preferred policy, to prescribe or not prescribe a preferred policy, to accept or reject a prediction about an expected outcome, to formulate a problem in one way rather than another.

In policy analysis, these methods have special names:

- *Problem structuring.* Problem-structuring methods are employed to produce knowledge about what problem to solve. Problem-structuring methods include the influence diagram and decision tree presented in Case 1.2 of this chapter ("Using Influence Diagrams and Decision Trees to Structure Problems of Energy and Highway Safety Policy"). Other examples of problem-structuring methods include argument mapping (Case 1.3, "Mapping International Security and Energy Crises"). Chapter 3 of this book covers problem-structuring methods more extensively.
- *Forecasting.* Forecasting methods are used to produce knowledge about expected policy outcomes. Although many kinds of forecasting methods are covered in Chapter 4, an example of a simple forecasting tool is the scorecard described in Case 1.1 (The Goeller Scorecard and Technological Change). Scorecards, which are based on the judgments of experts, are useful in identifying expected outcomes of science and technology policies.
- *Prescription.* Methods of prescription are employed to create knowledge about preferred policies. An example of a prescriptive method is the spreadsheet (in Case 1.2, "Using Influence Diagrams and Decision Trees"). The spreadsheet goes beyond the identification of expected policy outcomes by expressing consequences in terms of monetary benefits and costs. Benefit–cost analysis and other methods of prescription are presented in Chapter 5.
- *Monitoring.* Methods of monitoring are employed to produce knowledge about observed policy outcomes. The scorecard (Case 1.1) is a simple method for monitoring observed policy outcomes as well as forecasting expected policy outcomes. Chapter 6 covers methods of monitoring in detail.
- *Evaluation.* Evaluation methods are used to produce knowledge about the value or utility of observed policy outcomes and their contributions to policy performance.

[12]Type I and Type II errors are also known as false positives and false negatives. Other sources on Type III errors include A. W. Kimball, "Errors of the Third Kind in Statistical Consulting," *Journal of the American Statistical Association* 52 (1957): 133–142; Howard Raiffa, *Decision Analysis* (Reading, MA: Addison-Wesley, 1968); and Ian I. Mitroff, *The Subjective Side of Science* (New York: Elsevier, 1974).

[13]John O'Shaughnessy, *Inquiry and Decision* (London: George Allen & Unwin, 1972).

Although evaluation methods are covered more fully in Chapter 7, the spreadsheet in Case 1.2 may be used for evaluation as well as prescription.

The first method, problem structuring, is about the other methods. For this reason, it is a *metamethod* (method of methods). In the course of structuring a problem, analysts typically experience a "troubled, perplexed, trying situation, where the difficulty is, as it were, spread throughout the entire situation, infecting it as a whole."[14] *Problem situations* are not problems, because problems are representations or models of problem situations. Hence, problems are not "out there" in the world but stem from the interaction of the thoughts of many persons and the external environments in which they work or live. It is important to understand that analysts with different perspectives see the same problem situation in different ways. Imagine a graph showing increased national defense expenditures in trillions of dollars. The problem situation, represented by the graph, may be seen by one stakeholder as evidence of increasing national security (more of the budget is allocated to defense) and by another as an indication of declining resources for social services (less of the budget can be allocated to social services). Problem structuring, a process of testing different formulations of a problem situation, governs the production and transformation of knowledge produced by other methods. Problem structuring, which is important for achieving approximate solutions to ill-structured or "wicked" problems,[15] is the central guidance system of policy analysis.

Policy-analytic methods are interdependent. It is not possible to use one method without using others. Thus, although it is possible to monitor past policies without forecasting their future consequences, it is usually not possible to forecast policies without first monitoring them.[16] Similarly, analysts can monitor policy outcomes without evaluating them, but it is not possible to evaluate an outcome without first monitoring the existence and magnitude of an outcome. Finally, to select a preferred policy typically requires that analysts have already monitored, evaluated, and forecasted outcomes.[17] This is another way of saying that policy prescription is based on factual as well as value premises.

Figure 1.1 supplies a framework for integrating methods from different disciplines and professions. The five policy-analytic methods are used across political science, sociology, economics, operations research, public administration, program evaluation, and ethics.

[14]John Dewey, *How We Think* (Boston, MA: D. C. Heath and Company, 1933), p. 108. The original statement of the difference between a problem and a problem situation is attributable to philosophical pragmatists including Charles Sanders Peirce and John Dewey.

[15]The original statement about "wicked" problems, a term coined by C. West Churchman in 1967, is Horst W. Rittel, "On the Planning Crisis: Systems Analysis of the First and Second Generations," pp. 390–396. Berkeley, CA: Institute of Urban and Regional Development. A more recent statement is Sandra S. Batie, "Wicked problems and applied economics." *American Journal of Agricultural Economics,* 90, 5 (2008): 1176–1191.

[16]An exception is predictions made on the basis of expert judgment or on abductive (retroductive) reasoning. The explanation of a policy is not necessary for predicting its future consequences. Strictly speaking, a prediction is a causal inference, whereas a projection, extrapolation, or "rational forecast" is not. On abductive reasoning in economics see Daniel W. Bromley, *Volitional Pragmatism and the Meaning of Economic Institutions* (Princeton, NJ: Princeton University Press, 2006).

[17]Causation may be assumed but not understood. Recipes claim only that a desired result is a consequence of action. Joseph L. Bower, "Descriptive Decision Theory from the 'Administrative' Viewpoint," in *The Study of Policy Formation,* ed. Raymond A. Bauer and Kenneth J. Gergen (New York: Free Press, 1968), p. 10.

Some methods are used solely or primarily in some disciplines, and not others. Program evaluation, for example, employs monitoring to investigate whether a policy is causally relevant to an observed policy outcome. Although program evaluation has made extensive use of interrupted time-series analysis, regression-discontinuity analysis, causal modeling, and other techniques associated with the design and analysis of field experiments,[18] implementation research within political science has not. Instead, implementation researchers have relied mainly on techniques of case study analysis.[19] Another example comes from forecasting. Although forecasting is central to both economics and systems analysis, economics has drawn almost exclusively on econometric techniques. Systems analysis has made greater use of qualitative forecasting techniques for synthesizing expert judgment, for example, Delphi technique and other qualitative techniques of synthesizing expert judgments.[20]

FORMS OF POLICY ANALYSIS

Relationships among types of knowledge and methods provide a basis for contrasting different forms of policy analysis (Figure 1.2).

Prospective and Retrospective Analysis

Prospective policy analysis involves the production and transformation of knowledge before prescriptions are made. Prospective, or *ex ante* analysis, shown as the right half of Figure 1.2, typifies the operating styles of economists, systems analysts, operations researchers, and decision analysts. The prospective form of analysis is what Williams means by policy *analysis.* Policy analysis is a means of analyzing knowledge "to draw from it policy alternatives and preferences stated in comparable, predicted quantitative and qualitative terms . . . it does not include the *gathering* of knowledge."[21] Policy *research,* by contrast, refers to studies that use social science methods to describe or explain phenomena. However, prospective analysis is often attended by gaps between the choice of preferred solutions and efforts to implement them. The reason for these gaps is that only a small amount of work required to achieve a desired set of objectives is carried out *before* policies are implemented. "It is not that we have too many good analytic solutions to problems," writes Graham Allison. "It is, rather, that we have more good solutions than we have appropriate actions."[22] *Retrospective policy analysis,* a potential solution, is displayed as the left half of Figure 1.2. This form of *ex post* analysis involves the production and transformation of knowledge *after* policies have been

[18]See, e.g., William R. Shadish, Thomas D. Cook, and Donald T. Campbell, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference* (Belmont, CA: Wadsworth, Cengage Learning, 2002).
[19]Paul A. Sabatier and Hank C. Jenkins-Smith, "The Advocacy Coalition Framework: An Assessment," in *Theories of the Policy Process,* ed. P. A. Sabatier (Boulder, CO: Westview Press, 1999), pp. 117–166.
[20]See Chapter 5.
[21]Walter Williams, *Social Policy Research and Analysis: The Experience in the Federal Social Agencies* (New York: American Elsevier, 1971), p. 8.
[22]Graham T. Allison, *Essence of Decision: Explaining the Cuban Missile Crisis* (Boston, MA: Little, Brown, 1971), pp. 267–268.

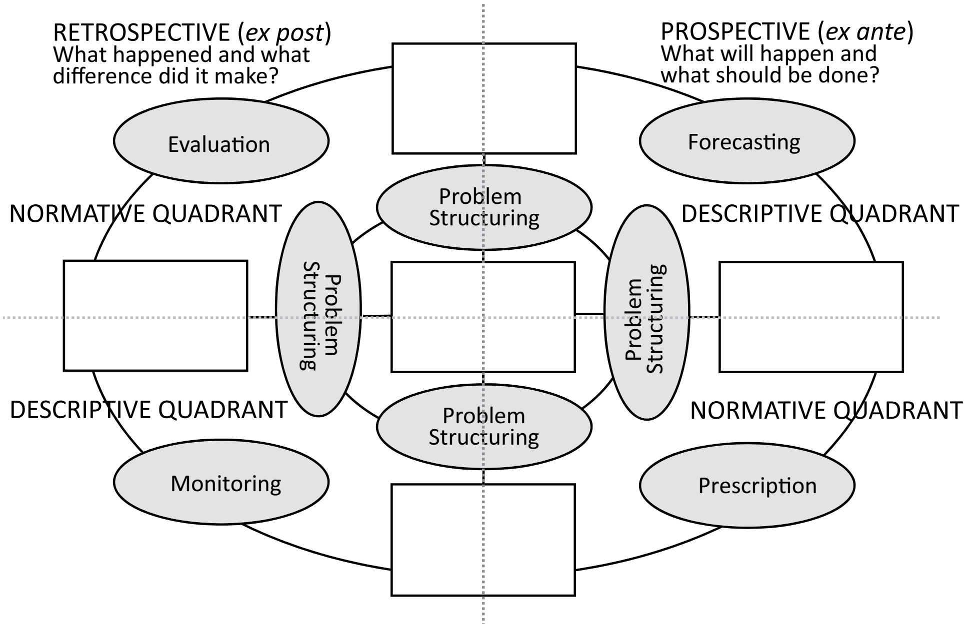


FIGURE 1.2
Forms of Policy Analysis

implemented. Retrospective analysis characterizes the operating styles of several groups of analysts:

- *Discipline-oriented analysts.* This group, comprised mainly of political scientists, economists, and sociologists, seeks to develop and test discipline-based theories about the causes and consequences of policies. This group is not concerned with the identification of "policy" variables that are subject to manipulation and those that are not.[23] For example, the analysis of the effects of party competition on government expenditures provides little or no knowledge about specific policy goals and how to attain them, because party competition is not a variable that policymakers can manipulate to change government expenditures.
- *Problem-oriented analysts.* This group, again composed mainly of political scientists, economists, and sociologists, also seeks to describe the causes and consequences of policies. However, problem-oriented analysts are less concerned, if at all, with the development and testing of theories that are important to social science disciplines. Rather they are concerned with identifying variables that may explain a problem. Problem-oriented analysts are not concerned with specific objectives, because the

[23]James S. Coleman, "Problems of Conceptualization and Measurement in Studying Policy Impacts," in *Public Policy Evaluation*, ed. Kenneth M. Dolbeare (Beverly Hills and London: Sage Publications, 1975), p. 25.

orientation of problem-oriented analysts is general, not specific. For example, the analysis of quantitative data on the effects of ethnicity and poverty on achievement test scores helps explain the causes of inadequate test performance, but does not provide knowledge about specific policies that can be manipulated to solve the problem.

- *Applications-oriented analysts.* A third group includes groups of applied political scientists, applied economists, applied sociologists, and applied psychologists, as well as members of professions such as public administration, social work, and evaluation research. This group also seeks to describe the causes and consequences of policies and pays little attention to the development and testing of theories. However, this group, in contrast to the previously described groups, is dedicated to the identification of manipulable policy variables that can potentially achieve specific objectives that can be monitored and evaluated to evaluate the success of policies.[24] For example, applications-oriented analysts may address early childhood reading readiness programs that can be used to achieve higher scores on standardized tests.

The operating styles of the three groups reflect their characteristic strengths and weaknesses. Discipline-oriented and problem-oriented analysts seldom produce knowledge that is directly useful to policymakers. Even when problem-oriented analysts investigate problems such as educational opportunity, energy conservation, and crime control, the result is *macronegative* knowledge, that is, knowledge that describes the basic (or "root") causes of policies with the aim of showing why policies do *not* work. By contrast, *micropositive* knowledge shows what policies *do work* under specified conditions.[25] In this context, it is practically important to know that a specific form of gun control reduces the commission of crimes or that intensive police patrolling is a deterrent to robberies. But it is of little practical value to know that the crime rate is higher in urban than rural areas, or that there is an association between crime and poverty.

Even when applications-oriented analysts provide micropositive knowledge, however, they may find it difficult to communicate with practitioners of *ex ante* policy analysis, because applications-oriented analysts focus on observed (retrospective) outcomes rather than expected (prospective) ones. In agency settings, *ex ante* analysts, whose job it is to find optimally efficient future solutions, generally lack knowledge about policy outcomes observed through retrospective analysis. Practitioners of *ex ante* analysis also may fail to specify in sufficient detail the kinds of policy-relevant knowledge that will be most useful for monitoring, evaluating, and implementing their recommendations. For this reason, as Alice Rivlin has noted, "the intended outcomes of a policy are so vague that "almost any evaluation of it may be regarded as irrelevant because it missed the 'problem' toward which the policy was directed."[26] Moreover, legislators often formulate problems

---

[24]An important transition in economics is the movement towards applications-oriented economics in the form of "natural experiments." A concise history of this transition is Joshua D. Angrist and Jörn-Steffen Pischke, "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics," *Journal of Economic Perspectives* 24, 2 (Spring 2010): 3–30.

[25]Williams, *Social Policy Research and Analysis,* p. 8.

[26]Ibid. p. 13, and Alice Rivlin, *Systematic Thinking for Social Action* (Washington, DC: Brookings, 1971).

in general terms in order to gain acceptance, forestall opposition, and maintain political flexibility.

Contrasts among the operating styles of analysts suggest that discipline-oriented and problem-oriented analysis provides knowledge that is less useful than applications-oriented analysis, because retrospective (*ex post*) analysis as a whole is perhaps less effective in solving problems than prospective (*ex ante*) analysis. This conclusion, frequently argued by economists, has merit from the point of view that policymakers need advice on what actions to take in future. Nevertheless, retrospective analysis has important benefits. It places primary emphasis on the results of observed outcomes of action—that is, what has actually worked or not—and is not content with speculations about expected policy outcomes. Discipline-oriented and problem-oriented analysis may offer new frameworks for understanding policymaking, challenging conventional wisdom, challenging social and economic myths, and shaping the climate of opinion. Retrospective analysis, while it has affected intellectual priorities and understandings, has performed less well in offering potential solutions for specific problems.[27]

Descriptive and Normative Analysis

Figure 1.2 also captures another important contrast, the distinction between descriptive and normative analysis. Descriptive policy analysis parallels *descriptive decision theory,* which refers to a set of logically consistent propositions that describe or explain action.[28] Descriptive decision theories may be tested against observations obtained through monitoring and forecasting. Descriptive theories and conceptual frameworks tend to originate in political science, sociology, and economics. The main function of these theories and frameworks is to explain, understand, and predict policies by identifying patterns of causality, also known as *causal mechanisms.* The function of approaches to monitoring such as field experimentation is to establish the approximate validity of causal inferences relating policies to their presumed outcomes.[29] In Figure 1.2, descriptive policy analysis can be visualized as an axis moving from the lower left quadrant (monitoring) to the upper right quadrant (forecasting).

Normative policy analysis parallels *normative decision theory,* which refers to a set of logically consistent propositions that evaluate or prescribe action.[30] In Figure 1.2, normative policy analysis can be visualized as an axis running from the lower right quadrant (prescription) to the upper left quadrant (evaluation). Different kinds of knowledge are required to test normative, as distinguished from, descriptive decision theories. Methods of evaluation and prescription provide knowledge about policy performance and preferred policies, policies that have been or will be optimally efficient because benefits outweigh costs, or optimally equitable because those most in need are made better off. One of the most important

[27]Janet A. Weiss, "Using Social Science for Social Policy," *Policy Studies Journal* 4 (Spring 1976): 237.
[28]Bower, "Descriptive Decision Theory," p. 104.
[29]See Thomas D. Cook and Donald T. Campbell, *Quasi-Experimentation: Design and Analysis Issues for Field Settings* (Boston, MA: Houghton Mifflin, 1979); Shadish, Cook, and Campbell, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.*
[30]Bower, "Descriptive Decision Theory," pp. 104–105.

features of normative policy analysis is that its propositions rest on values such as efficiency, effectiveness, equity, responsiveness, liberty, enlightenment, and security.[31]

Problem Structuring and Problem Solving

The internal and external cycles of Figure 1.2 provide another important distinction. The inner cycle designates processes of *problem structuring*. Procedures of problem structuring are designed to identify elements that go into the definition of a problem, but not to identify solutions. What are the main elements of a problem? Are they political, economic, social, ethical, or all of these? How is the problem structured, that is, organized into a specific configuration of elements, for example a linear sequence or as a complex system? Who are the most important stakeholders who affect and are affected by the problem? Have appropriate objectives been identified? Which alternatives are available to achieve the objectives? Are there uncertain events that should be taken into account? Are we solving the "right" problem rather than the "wrong" one?

By contrast, problem-solving methods are located in the outer cycle of Figure 1.2. They are designed to solve rather than structure a problem. Problem solving is primarily technical in nature, in contrast to problem structuring, which is primarily conceptual. Methods such as econometric forecasting or benefit–cost analysis are problem-solving methods. For problems to be solved, they must be appropriately structured. For instance, problem-solving methods answer questions about the amount of variance in an outcome that may be explained by an independent variable. What is the probability of obtaining variance as large or larger than that obtained? What are the net benefits of different policies? What is their expected utility or payoff?

Integrated and Segmented Analysis

Integrated policy analysis links the two halves of Figure 1.2, the four quadrants, and the inner and outer cycles. Retrospective and prospective forms of analysis are joined in one continuous process. Descriptive and normative forms of analysis are also linked, as are methods designed to structure problems and also solve them. Practically speaking, this means that integrated policy analysis bridges the several main segments of multidisciplinary policy analysis, especially economics and political science.

Today, this need is not being properly met by specialized social science disciplines, which often produce *segmented policy analysis*. Today, the job of bridging these segmented disciplines is carried out within multidisciplinary professions including public administration, planning, management, and policy analysis. The American Society for Public Administration (ASPA), the National Association of Schools of Public Affairs and Administration (NASPAA), the American Planning Association (APA), the International Association of Schools and Institutes of Administration (IASIA), the Academy of Management (AM), the Operations Research Society of America (ORSA), and the Association for Public Policy and

[31]The classic treatise on these and other base-values is Harold. D. Lasswell and Abraham Kaplan, *Power and Society: A Framework for Political Inquiry* (New Haven, CT: Yale University Press, 1950).

Management (APPAM) are multidisciplinary professions. So far, these professions have been far more open to economics, political science, and other social sciences disciplines than these disciplines have been open to them, notwithstanding a virtual consensus among policy analysts and practitioners that multidisciplinary research is essential for solving real-world problems, which do not respect disciplinary boundaries.

The framework for integrated policy analysis with which we began this chapter (Figure 1.1) helps examine the assumptions, strengths, and limitations of methods employed in disciplines that are so specialized that they are difficult to employ for practical problem solving. However, the framework for integrated policy analysis identifies and relates the goals of methods of policy analysis, enabling us to see the specific functions performed by methods of problem structuring, monitoring, evaluation, forecasting, and prescription. The framework identifies different types of policy analysis: prospective (*ex ante*) and retrospective (*ex post*), descriptive and normative, and problem structuring and problem solving. The framework explains why we define policy analysis as a problem-solving discipline that links social science theories, methods, and substantive findings to solve practical problems.

THE PRACTICE OF POLICY ANALYSIS

Reconstructed Logic versus Logic-in-Use

The process of integrated policy analysis is a *logical reconstruction.*[32] The process of actually doing policy analysis never conforms exactly to this reconstruction, because all logical reconstructions are abstractions of behaviors and not literal descriptions of them.[33] This lack of conformity is captured by the term *logic-in-use,* which refers to the way practicing analysts actually do their work. Notably, that work departs significantly from the methodological "best practices" mandated by logical reconstructions.[34] Variation from best practices depends on a number of factors including the personal characteristics of analysts, their professional socialization, and the institutional settings in which they work.

- *Cognitive styles.* The personal cognitive styles of analysts predispose them toward different modes of acquiring, interpreting, and using knowledge.[35] Corporations, nonprofit organizations, and public agencies such as the U.S. Department of Corrections and the National Science Foundation use various tools to diagnose differences in cognitive styles to make optimal personnel placements.
- *Analytic roles.* In agency settings, most analysts are relatively insulated from politics. When this occurs, they may be described as "technicians." However, others perform roles that, in addition to technical content, are political. These "politicians"

[32]Also called *reconstructed logic.*

[33]On reconstructed logic and logic-in-use, see Kaplan, *The Conduct of Inquiry,* pp. 3–11.

[34]For a thoughtful analysis of this problem, see Hal. K. Colebatch, Robert Hoppe, and Mirko Noordegraaf, *Working for Policy* (Amsterdam: Amsterdam University Press, 2010).

[35]Studies using the Myers-Briggs type indicator (Jungian personality types) suggest different cognitive styles among scientists, managers, and analysts. References provided by the Myers and Briggs Foundation at www.myersbriggs.org. See also Ian I. Mitroff and Ralph H. Kilmann, *Methodological Approaches to Social Science* (San Francisco, CA: Jossey-Bass, 1978).

are actively committed to advancing the interests of political leaders or officials to whom they report, usually at higher levels of government. Other activist analysts are "entrepreneurs" who seek greater influence in the policymaking process.[36]

- *Institutional incentive systems.* Policy "think tanks" encourage different orientations toward analysis, including the "humanistic-value-critical" and the "scientific."[37] The institutional rewards and punishments provided by think tanks and other professional research organizations affect the validity of conclusions and recommendations.[38]
- *Institutional time constraints.* Analysts working in governmental settings are often subject to tight institutional time constraints (24 hours to 3 days is typical). They work with much greater speed, and perhaps greater efficiency, than analysts in academic settings. Understandably, government analysts rarely collect original data and most do not employ complex and time-consuming techniques.[39]
- *Professional socialization.* Different disciplines and professions socialize their members into different norms and values. Analyses of published papers suggest that analysts employ formal-quantitative as well as informal-narrative approaches, although sound policy recommendations sometimes require formal-quantitative procedures.[40]
- *Multidisciplinary teamwork.* Much analysis conducted in public agencies is carried out by multidisciplinary teams. Some members have primary responsibility for the different types of analysis displayed in the two halves and the quadrants of Figure 1.2. Economists and decision analysts are typically more qualified to perform prospective (*ex ante*) analysis, whereas team members trained in political science, sociology, and program evaluation are usually better at retrospective (*ex post*) analysis. The effectiveness of teams depends on everyone acquiring a basic understanding of analytic methods employed by others.

Methodological Opportunity Costs

Integrated analysis has opportunity costs. Given limited time and resources, it is difficult to conduct systematic economic, political, and organizational analyses simultaneously.

[36]Arnold Meltsner, *Policy Analysts in the Bureaucracy* (Berkeley: University of California Press, 1976); Robert A. Heineman, William T. Bluhm, Steven A. Peterson, and Edward N. Kearney, *The World of the Policy Analyst* (Chatham, NJ: Chatham House, 1990).

[37]Pamela Doty, "Values in Policy Research," in *Values, Ethics, and the Practice of Policy Analysis,* ed. William N. Dunn (Lexington, MA: D. C. Heath, 1983).

[38]Donald T. Campbell, "Guidelines for Monitoring the Scientific Competence of Preventive Intervention Research Centers: An Exercise in the Sociology of Scientific Validity," *Knowledge: Creation, Diffusion, Utilization* 8, 3 (1987): 389–430.

[39]See P. J. Cook and J. W. Vaupel, "What Policy Analysts Do: Three Research Styles," *Journal of Policy Analysis and Management* 4, 3 (1985): 427–428. Agencies such as the U.S. Bureau of Labor Statistics and the U.S. Congressional Budget Office are exceptions.

[40]An early but representative overview of approaches is Janet A. Schneider, Nancy J. Stevens, and Louis G. Tornatzky, "Policy Research and Analysis: An Empirical Profile, 1975–1980," *Policy Sciences* 15, 2 (1982): 99–114.

Multiple triangulation,[41] or what Cook calls *critical multiplism,*[42] responds to some of these inadequacies. Another methodological opportunity cost is attempting to follow, without sufficient time and resources, the doctrines and principles of philosophies of science such as logical positivism, constructivism, or pragmatism.[43] For example, positivism today appears as a one-sided methodology and epistemology claiming that statements about the policy must be logically and empirically verifiable, expressed in a formal (ideal) language such as mathematical statistics, and confirmed by means of statements that correspond to some notion of objective reality, rather than a reality constituted by subjective meaningful actions of policy stakeholders. Logical positivism, as Cook argues, was the dominant methodology of policy analysis and program evaluation during the era of President Lyndon Johnson's War on Poverty. The advantage of critical multiplism over logical positivism is that multiplism provides a better approximation of what is true by employing procedures that triangulate from a variety of perspectives on what is worth knowing and what is known about policies.[44]

A disadvantage of multiplism lies in its costs. Triangulation among multiple perspectives, along with the use of multiple methods, measures, and data sources, involves significant opportunity costs[45] (Figure 1.3). When single methods such as econometric modeling are employed to achieve measurement precision and statistical generalizability, analysts forgo opportunities to acquire a deeper understanding of policies that is possible through ethnographic interviews, case studies, and other qualitative methods. A leading econometrician, noting that economists are unique among social scientists because they are trained only to analyze existing data, not to collect it, observes that "empirical work can be greatly enhanced by being sensitive to the context of the problem (the data-generating process) and knowing a lot about one's data."[46] Similar tradeoffs apply to methods of research synthesis, or meta-analysis, which purchase measurement precision and generalized policy causation at the expense of a deeper understanding of contexts of policymaking.[47]

---

[41]The methodology of triangulation is analogous to practices employed in geodesic surveys, cartography, navigation, and, more recently, satellite tracking. The position or location of an object is found by means of bearings from two or more fixed points or electronic signals a known distance apart.

[42]Cook advanced critical multiplism as an alternative to logical positivism. See Thomas D. Cook, "Postpositivist Critical Multiplism," in *Social Science and Social Policy,* ed. R. Lane Shotland and Melvin M. Mark (Beverly Hills, CA: Sage Publications, 1985), pp. 21–62.

[43]A critical assessment of logical positivism is Mary E. Hawkesworth, "Epistemology and Policy Analysis," in *Advances in Policy Studies since 1950,* ed. Dunn and Kelly, pp. 293–328; and Mary E. Hawkesworth, *Theoretical Issues in Policy Analysis* (Albany: State University of New York Press, 1988).

[44]Cook, "Postpositivist Critical Multiplism," p. 57.

[45]See David Brinberg and Joseph E. McGrath, *Validity and the Research Process* (Beverly Hills, CA: Sage Publications, 1985). For Brinberg and McGrath and other methodological pragmatists, the choice of methods is similar to an optimization problem in decision analysis. See C. West Churchman, *Prediction and Optimal Decision: Philosophical Issues of a Science of Values* (Englewood Cliffs, NJ: Prentice Hall, 1961), and Russell Ackoff, *Scientific Method: Optimizing Applied Research Decisions* (New York: John Wiley, 1962).

[46]Peter Kennedy, *A Guide to Econometrics,* 4th ed. (Cambridge, MA: MIT Press, 1998), pp. 83–84.

[47]See Lawrence Rudner, Gene V. Glass, David L. Evartt, and Patrick J. Emery (2002), *A User's Guide to the Meta-Analysis of Research Studies,* ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD. http://echo.edres.org.

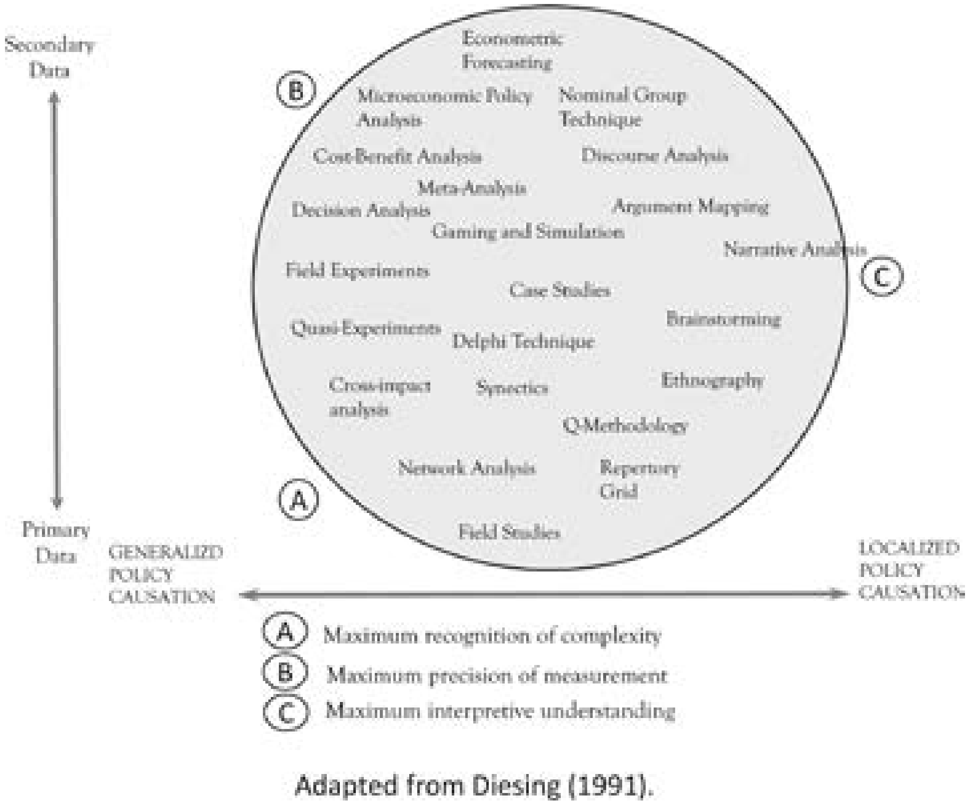


FIGURE 1.3
Opportunity Costs of Using Multiple Methods

Source: Adapted from Diesing (1991): 88.

Ethnographic interviews, by contrast, involve high knowledge costs because they require the collection of substantial primary data through interviews. However, they also lack precision and seldom permit the generalization of policy causation to other settings. Although greater precision can be obtained by means of field experiments, these are expensive, especially when they are employed in conjunction with mixed (quantitative and qualitative) methods. To be sure, triangulation among convergent (and divergent) perspectives, methods, and measures enhance the validity of policy analysis and other applied social sciences.[48] But the time and financial constraints make tradeoffs inevitable.

[48]The case for triangulation in its many forms is Donald T. Campbell, *Methodology and Epistemology for Social Science: Selected Papers,* ed. E. Samuel Overman (Chicago, IL: University of Chicago Press, 1988).

CRITICAL THINKING AND PUBLIC POLICY

Policy analysis is complex. Analysts must sift through large volumes of available data, evaluate sources of these data, select appropriate methods of analysis, and employ effective strategies for communicating the results. This challenge requires *critical thinking,* which involves the organization, synthesis, and evaluation of different reasons and bodies of evidence offered to support contending claims. One method available for this purpose is the analysis of *policy arguments.* By analyzing policy arguments, we can identify and probe the assumptions underlying competing policy claims, recognize and evaluate objections to these claims, and synthesize knowledge from different sources.

The Structure of Policy Arguments

Policy arguments are the main vehicle for carrying debates about public policies.[49] Although social scientists rightly pride themselves on the use of scientific methods, they too often forget that "public policy is made of language. Whether in written or oral form, argument is central to all stages of the policy process."[50]

The structure of a policy argument can be represented as a set of seven elements (Figure 1.4).[51]

Policy claim (C). A policy claim is the conclusion of a policy argument. Policy claims are of different types. Some are normative ("Congress should pass the amendments to the Fair Employment Practices Act") and some are descriptive ("The use of the Internet will double in the next 10 years"). Still others are evaluative ("The distribution of employment opportunities is unfair"), and some are definitional ("Unemployment is the number of persons in the active labor force seeking work").

Policy-relevant knowledge (K). Policy-relevant knowledge provides the grounds for a policy claim. These grounds may be statistical data, experimental findings, expert testimony, or common sense. Policy-relevant knowledge is a response to the question: What knowledge is relevant to the claim? Knowledge, which is the starting point of every argument, provides the grounds for making policy claims. Policy arguments may form argument chains and branch off into trees and cycles.

---

[49]See Frank Fischer and John Forester, eds., *The Argumentative Turn in Policy Analysis and Planning* (Durham, NC: Duke University Press, 1993). Earlier works are Ian I. Mitroff and Richard O. Mason, *Creating a Dialectical Social Science* (Boston, MA: D. Reidel, 1981); William N. Dunn, "Reforms as Arguments," *Science Communication* 3 (1982): 293–326; Des Gaspers and R. V. George, "Analyzing Argumentation in Planning and Public Policy: Improving and Transcending the Toulmin-Dunn Model," *Environment and Planning B: Planning and Design* 25 (1998): 367–390. Giandomenico Majone, *Evidence, Argument, and Persuasion in the Policy Process* (New Haven, CT: Yale University Press, 1989); and Deborah Stone, *Policy Paradox and Political Reason* (Glenview, IL: Scott Foresman, 1988).

[50]Majone, *Evidence, Argument, and Persuasion,* p. 1.

[51]This model of argument is built into the computer software called *Rationale 2,* which was developed by Tim van Gelder and colleagues in Australia (www.austhink.com). The classic structural model is presented in Stephen Toulmin, *The Uses of Argument* (Cambridge: Cambridge University Press, 1958).

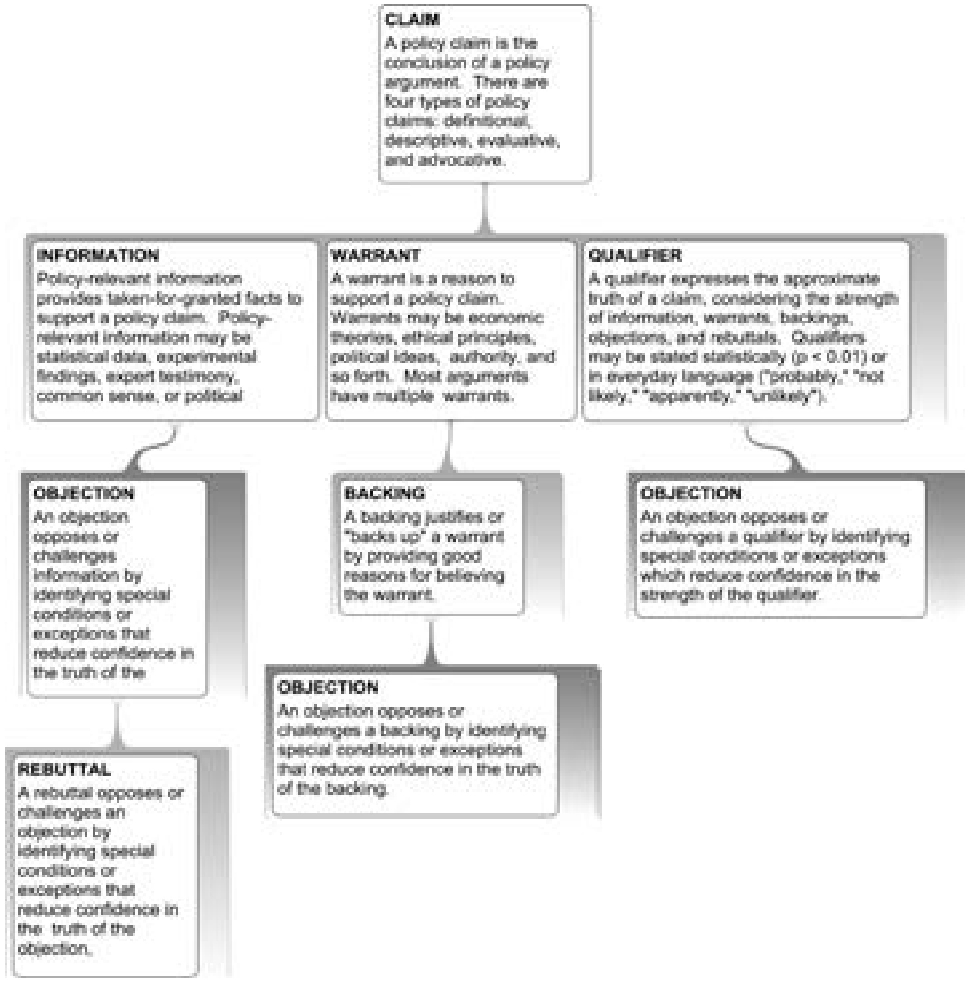


FIGURE 1.4
Elements of a Policy Argument

Warrant (W). The warrant is a reason to support a claim. Warrants may be economic theories, ethical principles, political ideas, or professional authority.[52] A warrant answers the question "Why does this reason support the claim? Different types of warrants are related to arguments made in different disciplines and professions. For example, law uses case comparisons and rules of evidence, while economics uses theories and component laws such as the law of diminishing utility of money. Policymakers as well as social scientists employ causal warrants such as "Ethnic cleansing will be deterred by air strikes that establish NATO's credibility in the region." The warrant, which provides a justification for accepting a claim, answers the question: *Considering the knowledge, what reasons make the claim true?*

[52]Different kinds of warrants yield the "modes" of policy argument discussed in Chapter 8.

Qualifier (Q). The qualifier expresses the degree to which a claim is approximately true, given the strength of the knowledge, warrants, backings, objections, and rebuttals. Although social scientists may state qualifiers in the language of formal probability ($p = 0.01$), ordinary language is the normal way of qualifying claims with such terms as: certainly, absolutely, probably, in all likelihood, presumably, barring unforeseen circumstances. The qualifier, which begins with the word *unless,* answers the question: *How strong or credible is the claim?* It is primarily through processes of argumentation and debate that policymakers, policy analysts, and other policy stakeholders adjust and sometimes abandon arguments. Such changes, when they occur, are based on the strength of objections and rebuttals.

Backing (B). The backing is an additional reason to support or "back up" the warrant. The backing, which begins with the word *because,* answers the question, *Why does the warrant support the claim?,* by providing a more general reason, assumption, or argument. Different kinds of backings are employed by different disciplines and professions. Backings may be scientific laws, rules of evidence, appeals to the authority of experts, or ethical and moral principles. For example, consider the warrant illustrated earlier: "Ethnic cleansing will be deterred by air strikes that establish NATO's credibility in the region." The backing for warrants advocating the use of coercive force is frequently an informal statement of the law of diminishing utility: "The greater the cost of an alternative, the less likely it will be pursued."

Objection (O). An objection opposes or challenges the knowledge, warrant, backing, or qualifier by identifying special conditions or exceptions that reduce confidence in the truth of the knowledge, warrant, backing, or qualifier. An objection, which begins with the word *but,* answers the question: *Are there special circumstances or exceptions that threaten the credibility of the warrant?* Analysts who incorporate objections into their analysis are more likely to adopt a critical perspective toward an issue by identifying weak or hidden assumptions, anticipating unintended consequences, or questioning possible rebuttals to objections. Thereby, analysts are self-critical, challenging their own assumptions and those of others.

Rebuttal (R). A rebuttal is an objection to an objection. Rebuttals oppose or challenge objections by identifying special conditions or exceptions that reduce confidence in the truth of the objection. Rebuttals, which begin with the word *however,* answer the question: *Are there special circumstances or exceptions that threaten the credibility of the warrant?* Many policy arguments have objections and rebuttals, because policymaking involves bargaining, negotiation, and competition among opponents and proponents of policies.

The perspectives of analysts may be found by identifying their underlying warrants, backings, objections, and rebuttals. Significantly, the identical policy-relevant knowledge is usually interpreted in distinctly different ways. A decrease in crime rates in urban areas may be welcomed by the urban poor, viewed with skepticism by owners of central city businesses, rejected by criminologists who attribute urban crime to changes in unemployment and homelessness, and hailed as an achievement by elected officials. By examining contending arguments and their underlying assumptions, analysts can uncover and critically assess reasons and evidence that otherwise may go unnoticed. Equally important is what critical thinking brings to analysts themselves, who are enabled to probe their own assumptions by examining the objections, qualifications, and exceptions to their own claims.

CHAPTER SUMMARY

This chapter has provided a methodology for policy analysis that identifies the functions of methods in creating and transforming policy-relevant knowledge. The four regions of the framework, along with inner and outer cycles, call attention to distinctive properties of these methods and point to their origins in different social science disciplines and professions. Because no one method is appropriate for all or most problems, there is a need for multidisciplinary analysis, which requires tradeoffs and opportunity costs. In turn, the analysis of tradeoffs and opportunity costs demands critical thinking. The analysis of policy arguments is well suited for this purpose.

REVIEW QUESTIONS

1. What does it mean to define policy analysis as a process of inquiry as distinguished from the application of methods of policy analysis?
2. Describe the dynamics of types of policy-relevant knowledge, policy-analytic methods, and knowledge transformations.
3. Contrast segmented and integrated policy analysis. Give examples.
4. How does normative decision theory differ from descriptive decision theory?
5. List some of the key differences between problem solving and problem structuring.
6. Contrast retrospective and prospective analysis. Which social science disciplines tend to specialize in prospective analysis? Retrospective analysis?
7. Discuss the strengths and limitations of critical multiplism.
8. Contrast the "logic-in-use" and the "reconstructed logic" of policy analysis. Provide examples.
9. How can argumentation mapping assist analysts to become critical thinkers?

DEMONSTRATION EXERCISES

1. Scorecards are a convenient way to represent complex outcomes and impacts of policies. At the same time, scorecards assume causal linkages that may not exist. What other methods are available to assess the effects of policies? Return to Figure 1.1 as a guide to your answer.
2. Influence diagrams and decision trees are useful for structuring policy problems. The diagram and tree displayed in Figure 1.4 help identify policy stakeholders, policy alternatives, uncertain outcomes and events, probabilities of these outcomes and events, and objectives. Consider the influence diagram, which represents the problem of energy supply in 1973 and 1974, when the OPEC oil embargo posed significant challenges to U.S. energy supply and national security. How does Figure 1.4 help us formulate the problem? What do the arrows suggest about the causes of the energy shortage as well as the causes of the decline in traffic fatalities? What does the influence diagram suggest about the conditions that gave rise to the 55 mph speed limit? After comparing the influence diagram with the decision tree (which is based on the influence diagram), describe why these two

problem representations are good examples of descriptive and normative decision theory.

3. Create an argument map based on the influence diagram presented in Case 1.3. Begin with the claim: "The United States should return to the 55 mph speed limit in order to conserve fuel and save lives." Include in your map as many warrants, backings, objections, and rebuttals as you can. Assuming that the original qualifier was "certainly," indicate whether the qualifier changes as we move from a simple, static, uncontested argument to a complex, dynamic, and contested one.

BIBLIOGRAPHY

Angrist, Joshua D., and Jörn-Steffen Pischke. "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con Out of Econometrics." *Journal of Economic Perspectives,* 24, 2 (Spring 2010): 3–30.

Campbell, Donald T. *Methodology and Epistemology for Social Science: Selected Papers.* Edited by E. Samuel Overman. Chicago, IL: University of Chicago Press, 1988.

Diesing, Paul. *How Social Science Works: Reflections on Practice.* Pittsburgh, PA: Pittsburgh University Press, 1991.

Dunn, William N., and Rita Mae Kelly. *Advances in Policy Studies since 1950.* New Brunswick, NJ: Transactions Books, 1992.

Fischer, Frank, and John Forester. *The Argumentative Turn in Policy Analysis and Planning.* Durham, NC: Duke University Press, 1993.

Imbens, Guido W., and Donald B. Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences.* Cambridge: Cambridge University Press, 2015.

Kaplan, Abraham. *The Conduct of Inquiry: Methodology for Behavioral Science.* San Francisco, CA: Chandler, 1964.

MacRae, Duncan, Jr. *The Social Function of Social Science.* New Haven, CT: Yale University Press, 1976.

Quade, Edward S. *Analysis for Public Decisions.* New York: American Elsevier, 1975.

Shadish, Will, Thomas D. Cook, and Donald T. Campbell. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.* Belmont, CA: Wadsworth, Cengage Learning, 2002.

Stone, Deborah. *Policy Paradox: The Art of Political Decision Making.* Rev. ed. New York: W. W. Norton, 2001.

Toulmin, Stephen E. *Return to Reason.* Cambridge, MA: Harvard University Press, 2001.

Toulmin, Stephen E., Richard Rieke, and Alan Janik, *An Introduction to Reasoning.* 2nd ed. New York: Macmillan, 1984.

Van Gelder, Tim. "The Rationale for Rationale." *Law, Probability, and Risk* 6 (2007): 23–42.

CASE 1.1 THE GOELLER SCORECARD AND TECHNOLOGICAL CHANGE

When advanced technologies are used to achieve policy goals, complex social and technical systems are involved. While it is analytically tempting to prepare a comprehensive economic analysis of the costs and benefits of such technologies, many practicing analysts do not have the time or the resources to do so. Many analyses are completed in a period of several days to a month, and in most cases policy analyses do not involve the collection and analysis of new data. Early on in a project policymakers and their staffs typically want an overview of the problem and the potential impacts of

(*continued*)

alternative policies. Under these circumstances, the scorecard is appropriate.

The Goeller *scorecard,* named after Bruce Goeller of the RAND Corporation, is appropriate for this purpose. Table C1.1 shows the impacts of alternative transportation systems. Some of the impacts involve transportation services used by members of the community, while others involve impacts on low-income groups. In this case, as Quade observes, the large number of diverse impacts are difficult to value in dollar terms, making a benefit–cost analysis impractical and even impossible.[53] Other impacts involve financial and economic questions such as investments, jobs created, sales, and tax revenues. Other impacts are distributional. ■

TABLE C1.1

Scorecard

Impacts	Conventional Takeoff and Landing Aircraft (CTOL)	Vertical Takeoff and Landing Aircraft (VTOL)	Tracked Air-Cushion Vehicle (TACV)
TRANSPORTATION			
Passengers (million miles)	7	4	9
Per trip time (hours)	2	1.5	2.5
Per trip cost ($)	$17	$28	$20
Reduced congestion (%)	0%	5%	10%
FINANCIAL			
Investment ($ millions)	$150	$200	$200
Annual subsidy ($ millions)	0	0	90
ECONOMIC			
Added jobs (thousands)	20	25	100
Added sales ($millions)	50	88	500
COMMUNITY			
Noise (households)	10	1	20
Added air pollution (%)	3%	9%	1%
Petroleum savings (%)	0%	−20%	30%
Displaced households	0	20	500
Taxes lost ($millions)	0	0.2	2
Landmarks destroyed	None	None	Fort X
DISTRIBUTIONAL			
Low-income trips (%)	7%	1%	20%
Low-income household			
Noise-annoyance (%)	2%	16%	40%

Source: Quade (1975): 60.

[53] E. S. Quade, *Analysis for Public Decisions* (New York: American Elsevier, 1975), p. 65.

CASE 1.2 USING INFLUENCE DIAGRAMS AND DECISION TREES TO STRUCTURE PROBLEMS OF ENERGY AND HIGHWAY SAFETY POLICY

Along with other methods of policy analysis discussed earlier in this chapter (Figure 1.1), the *influence diagram* and *decision tree* are useful tools for structuring policy problems.[54] The influence diagram (Figure C1.2.1) displays the policy, the National Maximum Speed Limit (NMSL) of 1974, which established the 55 mph speed limit. The rectangle in Figure C1.2.1 refers to a policy choice, or decision node, which in this case is the choice between adopting the national maximum speed limit of 55 mph and not adopting it. To the right and above the decision node are uncertain events, represented as ovals, which are connected to the decision node with arrows showing how the speed limit affects or is affected by the uncertain events. The rectangles with shaved corners represent valued policy outcomes, or objectives. The objectives are to lower fuel consumption, reduce travel time, and avert traffic fatalities and injuries. To the right of the objectives is another shaved rectangle, which designates the net benefits (benefits less costs) of the four objectives. In this case, the surprising result of using the influence diagram was the discovery of causally relevant economic events such as recession and unemployment, which affect miles driven, which in turn affects the achievement of all four objectives. The "root cause" appears to be the OPEC oil embargo.

The decision tree is another way to represent the influence diagram. Whereas the influence diagram shows how policy choices and uncertain events affect the achievement of objectives, the decision tree displays the monetary value of these outcomes. In this abridged and simplified decision tree, there are two branches that not only represent the two alternatives (renew or not renew the NMSL), but also the OPEC oil crisis, the recession, the costs of miles traveled, and the dollar benefits of reducing fatalities. The bolded branches show the events with the greatest likelihood of occurring, or which already have occurred. ■

(*continued*)

[54]The diagram and tree were created with the Decision Programming Language (DPL), which is available from Syncopation Software at www.syncopation.com. Educational, professional, and commercial versions of DPL 7.0 are available.

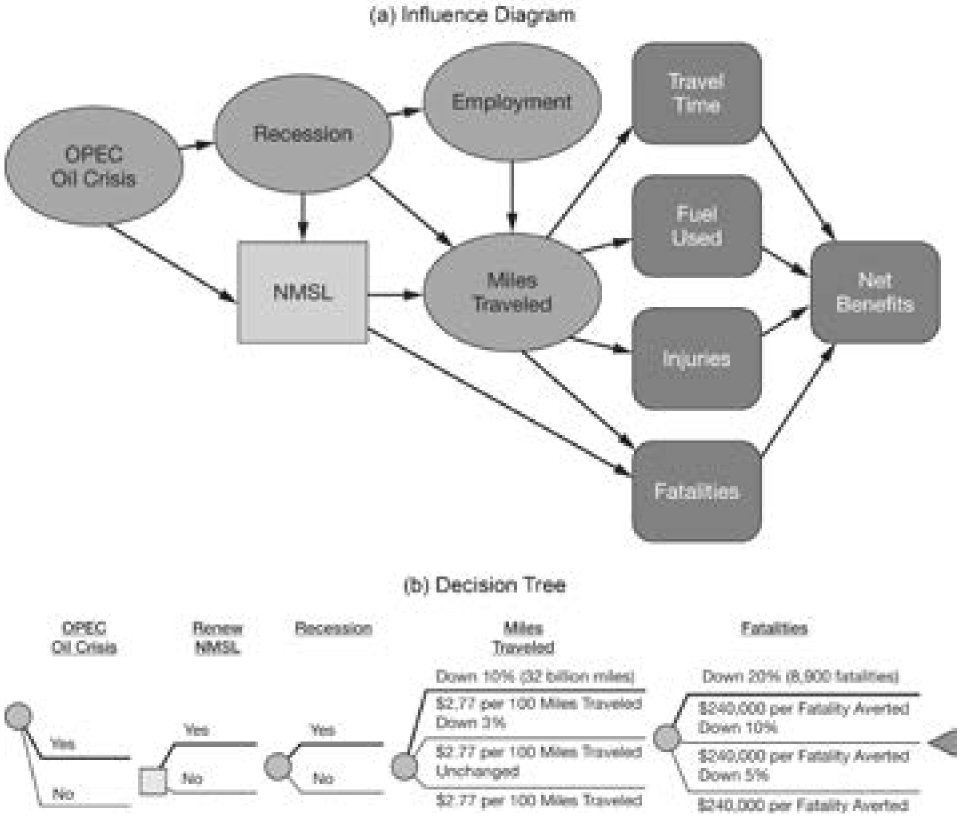


FIGURE C1.2.1
Influence Diagram and Decision Tree

CASE 1.3 MAPPING INTERNATIONAL SECURITY AND ENERGY CRISES

Another way to represent problems is to use the argument mapping methods introduced earlier in this chapter (Figure 1.3). The role of argument mapping in structuring policy discourses may be illustrated by Graham Allison's well-known study of foreign policy decision-making during the Cuban missile crisis of October 1962.[55] Showing how different explanatory models yield different conclusions, Allison argues that government policy analysts use implicit conceptual models to think about and make policy recommendations. These conceptual models explain the behavior of governments in terms that assume the rationality of political choices (rational actor model), the inertia created by organizational processes (organizational process model), and the effects of bureaucratic politics (bureaucratic politics model). Each of these models provides its own explanation of policymaking.

In 1962, the policy alternatives open to the United States in responding to the crisis ranged from no action and diplomatic pressure to secret negotiations, invasion, surgical air strikes, and blockade. Among the several claims made at the time of the Cuban missile crisis, let us consider the policy actually adopted by the United States: "The United States should blockade Cuba." In this case, the policy-relevant knowledge (I) is "The Soviet Union is placing offensive missiles in Cuba." The warrant states that "the blockade will force the withdrawal of missiles by showing the Russians that the United States is determined to use force." In providing reasons to accept the warrant, the backing (B) supports the warrant by stating that "an increase in the cost of an alternative reduces the likelihood of that alternative being chosen."[56] The backing (B) represents a general theoretical proposition, or law, within the rational policy model. After the objection (O) has successfully challenged the warrant, the qualifier (Q) changes from *absolutely* to *doubtful*.

Allison's account shows how the use of multiple competing explanations can facilitate critical thinking. The use of multiple competing models moves the analysis from a simple uncontested argument (Figure C1.3.1) to a new argument which is complex, contested, and dynamic (Figure C1.3.2). This change occurs because a serious objection has been raised about the warrant and the backing of the claim. The objection states: "But Soviet leaders may fail to convince their naval units to depart from established organizational routines." The warrant for this objection is: "The bulk of research on organizations shows that major lines of organizational behavior tend to be straight. Behavior at time $t + 1$ differs little from behavior at time t.[57] The blockade will not work." The warrant for the objection is again a general proposition or law within the organizational process model, otherwise known as the disjointed incremental theory of policy change.

Simple uncontested maps of arguments about the OPEC oil embargo are presented alongside arguments about the Cuban missile crisis (Figures C1.3.1[a] and [b]). The comparisons show that the simple argument maps tend to result in uncritical thinking. By contrast, the complex, dynamic, and contested maps of the same crises (Figures C1.3.2[a] and [b]) illustrate what is meant by *critical thinking*. ■

(*continued*)

[55]Graham T. Allison, "Conceptual Models and the Cuban Missile Crisis," *American Political Science Review* 3002, 3 (1969): 689–718.
[56]Allison, "Conceptual Models," p. 694.
[57]Ibid., p. 702.

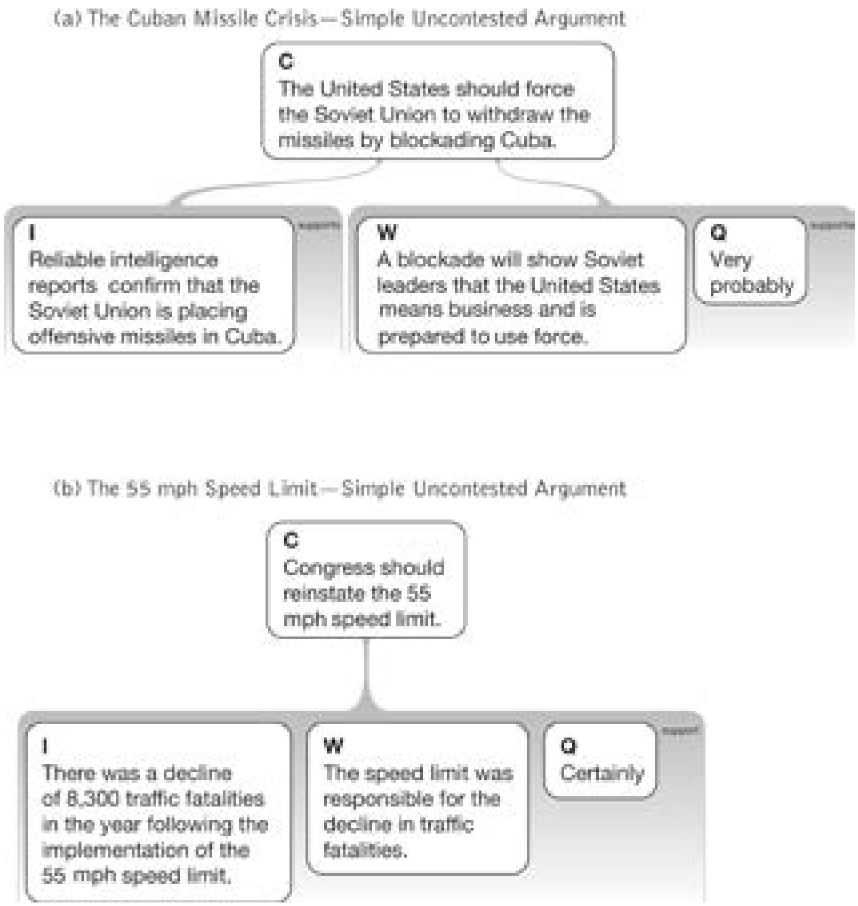
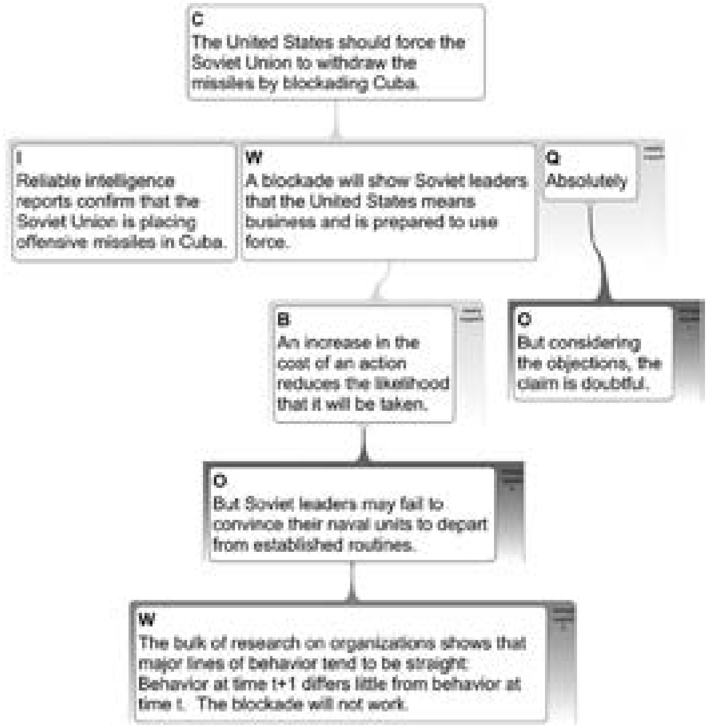


FIGURE C1.3.1
Simple Argument Maps Are Static and Uncontested

(a) The Cuban Missile Crisis—Dynamic Contested Argument



(b) The 55 mph Speed Limit—Simple Uncontested Argument

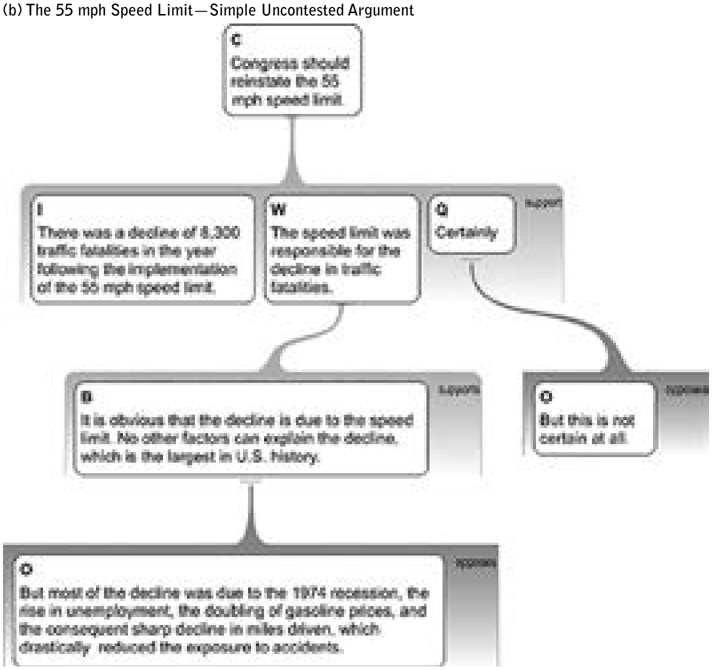


FIGURE C1.3.2
Complex Argument Maps Are Dynamic and Contested

CHAPTER 2

Policy Analysis in the Policymaking Process

LEARNING OBJECTIVES

By studying this chapter, you should be able to:

- Understand policy analysis as an intellectual activity embedded in a social process
- Explain the historical development of policy analysis as a response to practical problems
- Describe policymaking as a complex process with multiple phases ordered in time
- Contrast competing explanations of policy continuity and change
- Identify potential and actual uses of policy analysis
- Distinguish the composition, scope, and expected effects of using knowledge
- Use criteria associated with "truth" and "utility" to explain the uses of analysis

INTRODUCTION

Policy analysts create, critically assess, and communicate knowledge about and in the policymaking process.[1] The distinction between *about* and *in* marks an essential difference between policy analysis, on one hand, and political science and economics, disciplines that specialize in developing and testing empirically theories about policymaking. Although some members of these disciplines work on practical problems facing policymakers, the majority are motivated by incentive systems that demand the production of knowledge for its own sake. By contrast, policy analysts work under incentives designed to promote practical knowledge—typically, knowledge of what works.[2] Although policy analysis draws on traditional social science methods, theories, and substantive findings, the aim is to solve practical problems. This requires the communication and use of knowledge *in* as well as *about* the policymaking process.[3]

THE HISTORICAL CONTEXT

In a broad sense, policy analysis is as old as civilization itself. It includes diverse forms of inquiry, from mysticism and the occult to modern science. Etymologically, the term *policy* comes from the Greek, Sanskrit, and Latin languages. The Greek *polis* (city-state) and Sanskrit *pur* (city) evolved into the Latin *politia* (state) and later into the Middle English *policie*. The latter referred to the conduct of public affairs or the administration of the state.

The etymological origins of the word policy are the same for two other important words: *police* and *politics*. These multiple connotations are found in Germanic and Slavic languages, which have only one word (*Politik, politika*) to refer to both policy and politics. This is among the reasons for the porous boundaries among political science, public administration, and policy analysis, and the resultant confusion about their aims.

[1]Harold D. Lasswell, *A Pre-view of Policy Sciences* (New York: American Elsevier Publishing, 1971), pp. 1–2. For Lasswell, knowledge *about* refers to "systematic, empirical studies of how policies are made and put into effect," while knowledge *in* refers to the fact that "the realism of a decision depends in part on access to the stock of available knowledge."

[2]Here as elsewhere the term *knowledge* is used in preference to *information*. As Fritz Machlup shows, knowledge does not have to be scientific or intellectual to qualify as knowledge. Other types of knowledge are practical, pastime, religious, and unwanted. Fritz Machlup, *Knowledge: Its Creation, Distribution and Economic Significance*, Volume I: *Knowledge and Knowledge Production* (Princeton, NJ: Princeton University Press, 2014).

[3]Carol H. Weiss's *Social Science Research and Decision Making* (New York: Columbia University Press, 1980) is a comprehensive synthesis of research and theory on the uses of social science research in policymaking. Similar efforts include William N. Dunn and Burkart Holzner, "Knowledge in Society: Anatomy of an Emerging Field," *Knowledge in Society* (later titled *Knowledge and Policy*) 1, 1 (1988): 1–26; David J. Webber, "The Distribution and Use of Policy Knowledge in the Policy Process," in *Advances in Policy Studies since 1950*, ed. Dunn and Kelly (New Brunswick, NJ: Transaction, 1991), pp. 415–441; and National Academy of Sciences, *Using Science as Evidence in Public Policy* (Washington, DC: National Academies Press, 2009). Journals that focus on these problems include *Science Communication* (formerly *Knowledge: Creation, Diffusion, Utilization*) and *Science, Technology and Human Values.*

EARLY ORIGINS

The term *policy analysis* need not be restricted to its contemporary meaning, where analysis refers to breaking problems into basic elements or parts, much as we disassemble a clock or a vehicle. This is the sense of "analysis" when the term is used to speak about the decomposition of a decision into options, alternatives, and outcomes. A related view is that policy analysis is a collection of quantitative techniques used by systems analysts, decision analysts, and economists to examine the likelihood and utility of policy outcomes.

Understood in a wider sense, however, policy analysis may be seen to have emerged at some point in the evolution of human societies where practical knowledge was consciously cultivated, thereby prompting a self-reflective examination of links between knowledge and action. The development of specialized procedures for analyzing policies was related to the emergence of urban civilization out of scattered and largely autonomous tribal and folk societies.[4] As a specialized activity, policy analysis followed changes in social and, above all, political organization that accompanied new production technologies and stable patterns of human settlement.

An early effort to cultivate policy-relevant knowledge occurred in Mesopotamia, in what is now the Basra region of southern Iraq. In the ancient Mesopotamian city of Ur, one of the first legal codes was produced in the twenty-first century B.C., some two thousand years before Aristotle (384–322 B.C.), Confucius (551–479 B.C.), and Kautilya (ca. 300 B.C.) produced classic treatises on government and politics. In the eighteenth century B.C., the ruler of Babylon, with the assistance of professionals whom we would now call policy analysts, created the Code of Hammurabi. The Code was designed to establish a unified and just public order in a period when Babylon was in transition from a small city-state to a large territorial state. The Hammurabian Code, a set of policies that parallel Mosaic laws, reflected the economic and social requirements of stable urban settlements where rights and obligations were defined according to social position. The Code covered criminal procedures, property rights, trade and commerce, family and marital relations, physicians' fees, and what we now call public accountability.[5]

The early Mesopotamian legal codes were a response to the growing complexity of fixed urban settlements, where policies were needed to regulate the distribution of commodities and services, organize record keeping, and maintain internal security and external defense. A growing consciousness of relations between knowledge and action fostered the growth of educated strata that specialized in the production of policy-relevant knowledge. These "symbol specialists," as Lasswell called them, were responsible for policy forecasting. They were expected to forecast crop yields at the onset of the planting season and predict the outcomes of war.[6]

Because analysts used mysticism, ritual, and the occult to forecast the future, their methods were difficult to replicate. To be sure, these and other procedures were based in part on evidence acquired through experience. However, any reasonable definition of

---

[4]Lasswell, *A Pre-view of Policy Sciences*, pp. 9, 13.

[5]*The Code of Hammurabi*, trans. Robert F. Harper (Chicago, IL: University of Chicago Press, 1904).

[6]Lasswell, *A Pre-view of Policy Sciences*, p. 11.

science requires that knowledge claims be assessed against observations that are independent of the hopes of analysts, or of those who hire them.[7]

Then as now, policy-relevant knowledge was ultimately judged according to its success (or failure) in shaping better policies, not simply because special methods were used to produce it. The ancients knew what some contemporary analysts forget—when methods are used to perform functions of ritualistic purification and symbolic legitimation, analysts and their clients eventually must face the issue of practical performance.[8] Statements such as "drug policy is based on good science," or that public opinion results are based on a "scientific" sample, are in vogue. But the invocation of "science" in this context may represent little more than symbolic legitimation or ritualistic purification.

In India, Kautilya's *Arthashastra,* written in the fourth century B.C., is a systematic guide to policymaking, statecraft, and government administration. The *Arthashastra* synthesized much that had been written up to that time on material success, or what we now call economics. Kautilya, an adviser to the Mauryan Empire in northern India, has been compared to Plato (427–327 B.C.), Aristotle (384–322 B.C.), and Machiavelli (1469–1527), all of whom were deeply involved in practical aspects of policymaking. Plato served as adviser to the rulers of Sicily, while Aristotle tutored Alexander of Macedonia from the time Alexander was 14 years old until he ascended the throne at the age of twenty. Although Aristotle, like some present-day social scientists, found practical politics repugnant, he seems to have accepted the assignment because he wanted to bring knowledge to bear on policy issues of the day. In this respect, he followed his teacher Plato, who said that good government would not occur until philosophers were kings, or kings philosophers. The opportunity to influence policy by instructing Alexander, the heir apparent, was an offer that in good conscience he could not refuse.[9]

These are examples of preeminent individual producers of specialized knowledge, not of groups of educated classes who later would influence policymaking in Europe and Asia. In the Middle Ages, the gradual expansion and differentiation of urban civilization brought with it an occupational structure that facilitated the development of specialized knowledge. Princes and kings recruited policy specialists to provide advice and technical assistance in areas where rulers were least able to make effective decisions: finance, war, and law. German sociologist Max Weber described the historical development of a class of educated policy specialists in these three areas: finance, law, and war. This "expert officialdom," wrote Weber, was triumphant by the sixteenth century.[10]

The growth of expert officialdom, also known as "professional politicians," assumed different forms in different parts of the world. In medieval Europe, India, China, Japan,

[7]Donald T. Campbell, "A Tribal Model of the Social System Vehicle Carrying Scientific Knowledge," in *Methodology and Epistemology for Social Science: Selected Papers,* ed. E. Samuel Overman (Chicago, IL: University of Chicago Press, 1988), pp. 489–503.

[8]Edward A. Suchman, "Action for What? A Critique of Evaluative Research," in *Evaluating Action Programs,* ed. Carol H. Weiss (Boston, MA: Allyn and Bacon, 1972), p. 81, and Martin Rein and Sheldon H. White, "Policy Research: Belief and Doubt," *Policy Analysis* 3, 2 (1977): 239–271.

[9]J.A.K. Thompson, *The Ethics of Aristotle: The Nichomachean Ethics Translated* (Baltimore, MD: Penguin Books. 1955), p. 11.

[10]Max Weber, "Politics as a Vocation," in *From Max Weber: Essays in Sociology*, ed. Hans C. Gerth and C. Wright Mills (New York: Oxford University Press, 1946), p. 88.

and Mongolia, the literate and technically useful clergy comprised expert officialdom. Christian, Brahmin, Buddhist, and Lamaist priests, much like many modern social scientists, earned a reputation for impartiality and disinterestedness insofar as they stood above practical politics and temptations of political power and economic gain. Educated men of letters—whose modern counterpart is the special presidential adviser—influenced policymaking until court nobles, who later came to dominate the political and diplomatic service, replaced them. In England, petty nobles and urban rentiers (investors) were recruited without compensation to manage local governments in their own interest. Jurists trained in Roman law and jurisprudence had a strong influence on policymaking, particularly in Continental Europe. They were largely responsible for the transformation of the late medieval state and the movement toward modern government.

The age of the Industrial Revolution was also that of the Enlightenment, a period in which a belief in human progress through science and technology became an ever more dominant theme among policymakers and their advisers. The development and testing of scientific theories of nature and society gradually became the preferred means for understanding and solving social problems. For the first time, policy-relevant knowledge was produced, for the most part, according to the canons of empiricism and the scientific method.

The Nineteenth-Century Transformation

In nineteenth-century Europe, producers of policy-relevant knowledge began to base their work on the systematic recording of empirical data. Earlier, philosophers and statesmen had offered explanations of policymaking and its role in society. Yet for several thousand years, there was an essential continuity in methods for investigating and solving social, economic, and political problems. If evidence for a particular point of view was provided, it was typically based on appeals to religious authority, ritual, or philosophical doctrine. What was new in the nineteenth century was a basic change in the procedures used to understand society and its problems, a change reflected in the growth of empirical, quantitative, and policy-oriented research.[11]

The first censuses were conducted in the United States (1790) and England (1801). It was at this time that statistics, sometimes known as "state arithmetic," and demography began to develop as specialized fields. The Manchester and London Statistical Societies, established in the 1830s, helped shape a new orientation toward policy-relevant knowledge. Organized by bankers, industrialists, and scholars, the societies hoped to replace traditional ways of thinking about social problems with empirical analyses of the effects of urbanization and unemployment on the lives of workers and their families. In the Manchester Society, an enthusiasm for quantification was coupled with a commitment to social reform, or "progress of social improvement in the manufacturing population."[12] The London Society, under the influence of Thomas Malthus (1766–1834), stressed that it confined its attention

[11]Daniel Lerner's essay, titled "Social Science: Whence and Whither?" treats this important transition. See *The Human Meaning of the Social Sciences*, ed. Daniel Lerner (New York: World Publishing, 1959), pp. 13–23.
[12]Nathan Glazer, "The Rise of Social Research in Europe," in *The Human Meaning of the Social Sciences*, ed. Daniel Lerner (New York: World Publishing, 1959), p. 51.

to facts, and, as far as possible, stated these facts numerically and in tables.[13] The London and Manchester societies used questionnaires to carry out studies, including paid "agents" who were the counterpart of today's professional interviewer or census enumerator. There were similar developments in France, Germany, and the Netherlands.

A preeminent contributor to the methodology of social statistics and survey research was Adolphe Quetelet (1796–1874), a Belgian mathematician and astronomer who was the major scientific adviser to the Dutch and Belgian governments.[14] Many topics now covered in present-day texts on survey design and analysis were addressed by Quetelet: questionnaire design; data collection, analysis, and interpretation; data organization and storage; and identification of conditions under which data are collected. In the same period, Frederic Le Play (1806–1882) wrote *Les Ouvriers Europeans* [*The European Workers*], a detailed empirical investigation of family income and expenditures of European workers in several countries. In Germany, Ernst Engel (1821–1896) sought to derive laws of "social economics" from empirical data expressed in statistical form.

In England, the work of Henry Mayhew and Charles Booth, who studied the life and employment conditions of the urban poor in natural (what we now call "field") settings, was representative of the new empirical approach to the study of social problems. Mayhew's *London Labour and the London Poor* (1851) described the lives of the laborers, peddlers, performers, and prostitutes who comprised London's urban underclass. In writing *Life and Labour of the People in London* (1891–1903), Booth employed school inspectors as key informants. Using what we now call *participant observation*, Booth lived among the urban poor, gaining firsthand experience of actual living conditions. A member of the Royal Commission on the Poor Law, he was an important influence on the revision of policies on old-age pensions. Booth's work also served as something of a model for policy-oriented research in the United States, including the *Hull House Maps and Papers* (1895) and W.E.B. Du Bois's *The Philadelphia Negro* (1899), both of which sought to document the scope and severity of poverty in urban areas.

The nineteenth-century transformation was not the result of declarations of methodological loyalty to canons of empiricism and the scientific method. Declarations to this effect did not and could not occur until the next century, when Vienna Circle philosophers engaged in the logical reconstruction of physics to propose formal principles and rules of scientific practice. The transformation came, it seems, from the uncertainty accompanying the shift from agrarian to industrial societies, a shift that actually preceded the Industrial Revolution, when industry and industrialized science required a politically stable order to operate efficiently. Political stability was a buffer against profound social instability.[15]

Thus, contrary to academic theorizing, science and technology was not responsible for the growth of newly centralized systems of political control. The situation was more complicated. Although science and technology contributed to problems of a newly uprooted, uneducated, and displaced class of urban workers and their families, it was the political steering of science and technology that was part of the solution. Dominant social groups

[13] Ibid., pp. 51–52.

[14] A significant history of statistics and statisticians is Stephen M. Stigler, *The History of Statistics: The Measurement of Uncertainty before 1900* (Cambridge, MA: Harvard University Press, 1990).

[15] J. H. Plumb, *The Growth of Political Stability in England, 1675–1725* (Baltimore, MD: Penguin Books, 1973), p. 12.

valued policy-oriented scientific research as a means to solidify and legitimate political and administrative control.

In the sphere of factory production, for example, the political organization of work preceded scientific and technological developments that later culminated in efficiency-enhancing machinery and the specialization of tasks.[16] In the sphere of public policy, we see a parallel development. Methods of empirical, quantitative, and policy-oriented research were a product of the recognition by bankers, industrialists, politicians, and the Victorian middle class that older methods for understanding the natural and social world were no longer adequate. The key questions of the day were practical: How much did members of the urban proletariat need to earn to maintain themselves and their families? How much did they have to earn before there was a taxable surplus? How much did they have to save from their earnings to pay for medical treatment and education? How much should capitalist owners and the state invest in day care facilities so that mothers might put in an effective day's work? How much investment in public works projects—sanitation, sewerage, housing, roads—was required to maintain a productive workforce and protect the middle and upper classes from infectious diseases cultivated in urban slums? Policy-oriented empirical research provided answers to these and other practical questions.

Twentieth-Century Professionalization

An important feature of the twentieth century, as compared with the nineteenth, is the professionalization of the social and behavioral sciences.[17] Twentieth-century producers of policy-relevant knowledge were no longer the heterogeneous group of bankers, industrialists, journalists, and university professors who coalesced around the early statistical societies. Rather, they were graduates with first and advanced degrees in policy-relevant disciplines and professions. Along with some of their professors, they occupied influential positions in government, working as consultants or researchers under contracts and grants. In background, experience, and qualifications, they were members of established professions that were guided by generally accepted scientific and professional standards.

The new professionals played an active role in the administration of Woodrow Wilson, particularly during World War I in the War Department. Later, under the Republican administration of Herbert Hoover, groups of social scientists carried out two major social surveys, *Recent Economic Trends* and *Recent Social Trends*. The largest influx of social scientists into government came, however, with Franklin Roosevelt's New Deal. Large numbers of social scientists staffed the numerous new agencies established during the Roosevelt administration (National Recovery Administration, Work Projects Administration, Public Works Administration, Securities and Exchange Commission, and Federal Housing Administration).

The primary function of social scientists in the 1930s was to investigate broad sets of potential solutions for general policy problems and, and not, as in later periods, to employ

[16]Stephen A. Marglin, "What Do Bosses Do? The Origins and Functions of Hierarchy in Capitalist Production," *Review of Radical Political Economy* 6, 2 (1974): 33–60.
[17]Elliot, Freidson, *Professional Powers: A Study of the Institutionalization of Formal Knowledge* (Chicago, IL: University of Chicago Press, 1986).

economic modeling, decision analysis, and policy experimentation to identify and select specific solutions to well-defined problems. The Roosevelt administration's National Planning Board (later the National Resources Planning Board), a majority of whose members were social and political scientists, illustrates the general approach to policy questions characteristic of the 1930s. The board was conceived as "a general staff, gathering and analyzing facts, observing the interrelation and administration of broad policies, proposing from time to time alternative lines of national procedure, based on thorough inquiry and mature consideration."[18] This same general orientation toward problems was evident among economists working for the Department of Agriculture, political scientists involved in the reorganization of the Executive Branch, and anthropologists conducting studies for the Bureau of Indian Affairs. However, social scientists also contributed to specific methodological innovations. For example, the Department of Agriculture led in developing the sample survey as a new research tool and instrument of government census policy.[19]

World War II and the period of postwar readjustment also provided social scientists with opportunities to demonstrate their value in solving practical problems. Interwar achievements in the area of survey research had laid a foundation for the use of interviews by the Office of War Information, the War Production Board, and the Office of Price Administration. Military and civilian agencies relied on social scientists to investigate problems of national security, social welfare, war production, price controls, and food rationing. The activities of agencies such as the Office of Strategic Services were continued after the war by the Office of Naval Research, by the Department of the Air Force, and, later, by the Research and Development Board (subsequently RAND) of the Department of Defense, and the Central Intelligence Agency. Special research institutes were established by the federal government, including the Operations Research Office of The Johns Hopkins University and the Human Resources Research Office at George Washington University. Among the seminal contributions to policy research in this period was *The American Soldier* (1950), a four-volume study produced by many notable social scientists. The director of the Army Morale Division commissioned this large-scale project in 1941, under the general direction of sociologist Samuel Stouffer. The project is significant, not only because of its scale but also because it was part of the development of governmental support for policy research and analysis. Military policymakers turned to the social researcher, not only for facts but also for causal inferences and conclusions that would affect the lives of millions of troops.[20] This large research program contributed to the development and refinement of multivariate statistical analysis and other quantitative techniques that are now widely used by researchers in all social science disciplines.

After World War II, the first systematic effort to develop an explicit policy orientation within the social and behavioral sciences was *The Policy Sciences: Recent Developments in Scope and Method* (1951), edited by political scientists Daniel Lerner and Harold

[18]Gene Lyons, *The Uneasy Partnership: Social Science and the Federal Government in the Twentieth Century* (New York: Russell Sage Foundation, 1969), p. 65.

[19]Harry Alpert, "The Growth of Social Research in the United States," in *The Human Meaning of the Social Sciences*, ed. Daniel Lerner (New York: World Publishing, 1959), pp. 79–80.

[20]Howard E. Freeman and Clarence C. Sherwood, *Social Research and Social Policy* (Englewood Cliffs, NJ: Prentice Hall, 1970), p. 25.

D. Lasswell. The "policy sciences," as Lasswell stated in the introduction, are not confined to theoretical aims of science; they also have a fundamentally practical orientation. Their purpose is not simply to provide a basis for making efficient decisions but also to provide knowledge needed to improve the democratic processes. "In a word, the special emphasis is upon the policy sciences of democracy, in which the ultimate goal is the practical realization of human dignity in theory and fact." Twenty years later, in 1971, Lasswell would acknowledge his debt to John Dewey and the founders of pragmatism. "The policy sciences," wrote Lasswell, "were a contemporary adaptation of the general approach to public policy recommended by John Dewey and other pragmatists."[21]

The systematic study of public policy also grew out of public administration, then a field within political science. In 1937, Harvard University established the Graduate School of Public Administration, which focused in part on public policy. In the late 1940s, an interuniversity committee was established to develop public policy curricula, the major product of which was Harold Stein's *Public Administration and Policy Development: A Case-Book* (1952). The interuniversity committee, composed of professors and practitioners of public administration, speaks for the close relationship between public policy and public administration before and after World War II.[22]

The impetus for developing methods and techniques of policy analysis—as distinguished from its theory and methodology—did not originate in political science or public administration. The technical side of policy analysis rather grew out of engineering, operations research, systems analysis, and applied mathematics. Many of those responsible for developing methods and techniques had received their formal training outside the social sciences. World War II had prompted the involvement of specialists whose orientation toward policy was primarily analytical rather than conceptual. The idea of "analysis" came to be associated with efforts to decompose problems into their fundamental components, for example, decomposing problems of national defense into policy alternatives (nuclear warheads, manned bombers, conventional ground troops) the consequences of which could be measured against the attainment of policy objectives. This new orientation, which budget specialist Alan Schick calls the "analycentric perspective,"[23] tended to preclude or restrict concerns with political and administrative aspects of public policy, for example, concerns with the political feasibility of alternatives or their implementation for purposes of democratic reform. Although Schick's "analycentric turn" represented a movement away from the contextual, problem-oriented, and multidisciplinary orientation of the policy sciences,[24] it has also supplied more systematic procedures for identifying, selecting, and estimating the benefits and costs of policies.[25]

The analycentric turn was accompanied by the growing influence of nonprofit research organizations ("think tanks") such as the RAND Corporation, which fostered the spread of

---

[21]Lasswell, *A Pre-view of Policy Sciences*, pp. xiii–xiv.

[22]H. George Frederickson and Charles Wise, *Public Administration and Public Policy* (Lexington, MA: D. C. Heath, 1977).

[23]Allen Schick, "Beyond Analysis," *Public Administration Review* 37, 3 (1977): 258–263.

[24]Peter de Leon, *Advice and Consent: The Development of the Policy Sciences* (New York: Russell Sage Foundation, 1988), Ch. 2.

[25]Martin Greenberger, Matthew A. Crenson, and Brian L. Crissey, *Models in the Policy Process: Public Decision Making in the Computer Era* (New York: Russell Sage Foundation, 1976), pp. 23–46.

systems analysis and other techniques to government agencies and the academic community.[26] The development of program-planning-budgeting systems (PPBS) was due in large measure to the efforts of operations researchers and economists working under Charles Hitch at the RAND Corporation. The RAND group wanted to find out how the United States could "purchase' national security in the most efficient manner: "how much of the national wealth should be devoted to defense, how the funds allocated to defense should be distributed among different military functions, and how to assure the most effective use of these funds."[27] Although PPBS was introduced into the Department of Defense in 1965 and later mandated for use in all federal agencies, it proved difficult to implement. After 1971 it became discretionary and soon fell into disuse. Despite mixed conclusions about its success as a tool of policy analysis, PPBS does appear to have captured the attention of government and university analysts who value systematic procedures for selecting and evaluating policy alternatives.[28]

The analycentric turn has been offset to some extent by the rapid growth of private foundations whose mission is to support traditional lines of research in the social sciences and humanities. More than three-fourths of these foundations were established after 1950.[29] In the period after 1950, the federal government began to set aside funds for policy-related research in the social sciences, although the natural sciences continued to receive the bulk of total government research support. Although funding for applied and basic research in the social sciences fell by approximately 40 percent in constant dollars in 1980–1990,[30] it is important to recognize that more than 95 percent of all research funded by all types of organizations—governmental, nonprofit, and private—is represented, not by basic research, but by applied research on practical problems.

By the 1970s, many social science disciplines had established institutions expressly committed to applied and policy-related research, each with its own journal of record. These include the Policy Studies Organization (Political Science), the Society for the Study of Social Problems (Sociology), and the Society for the Psychological Study of Social Issues (Psychology). In the 1980s, the process of institutionalizing policy-relevant social science was carried a step further by the creation of multidisciplinary professional associations such as the Association for Public Policy and Management, which holds annual research conferences and publishes a journal of record, the *Journal of Policy Analysis and Management*. The new journal brought an applied and practical focus, as distinguished from the predominantly intellectual focus of mainstream disciplinary journals such as *The American Political Science Review*, *The American Economic Review*, and *The American Sociological*

---

[26]Bruce L. R. Smith, *The Rand Corporation: A Case Study of a Nonprofit Advisory Corporation* (Cambridge, MA: Harvard University Press, 1966).

[27]Greenberger, Crenson, and Crissey, *Models in the Policy Process,* p. 32.

[28]For example, Alice Rivlin, *Systematic Thinking for Social Action* (Washington, DC: Brookings Institution, 1971), and Walter Williams, *Social Policy Research and Analysis: The Experience in the Federal Social Agencies* (New York: American Elsevier Publishing, 1971).

[29]Irving Louis Horowitz and James E. Katz, *Social Science and Public Policy in the United States* (New York: Praeger Publishers, 1975), p. 17.

[30]National Science Foundation, *Federal Funds for Research, Development, and Other Scientific Activities* (Washington, DC: NSF, 1973), and National Science Board, *Science and Engineering Indicators—1989* (Washington, DC: NSB, 1989).

Review. By contrast, journals such as *Policy Sciences,* the *Policy Studies Journal,* and *Policy Studies Review* helped create a more applied and policy-relevant orientation, as indicated partly by the term *policy* in their titles. In addition to mainstream journals were literally hundreds of other publications that focused on specific issues involving health, welfare, education, criminal justice, education, science and technology, national security, and other policy issue areas.[31]

In the same period, universities in the United States and Europe founded new graduate programs and degrees in policy analysis. In the United States, a number of new programs were established with the support of the Ford Foundation's Program in Public Policy and Social Organization. Most research universities in the United States have policy centers or institutes listed in the *Encyclopedia of Associations,* along with thousands of freestanding nonprofit policy research organizations and advocacy groups. Most were established after 1950. In Washington and most state capitals, and in the European Union, "policy analyst" is a formal job description. The National Governor's Association and the National League of Cities have policy analysis units. There are similar units throughout the United States government listed in directorates of the European Union, the United Nations, and the World Bank. Wikipedia lists tens of thousands of policy think tanks in more than 150 countries, while the *Encyclopedia of Organizations,* maintained by Gale Publishing Company, lists 23,000 organizations in the United States engaged in policy research or policy advocacy.[32]

The Era of Evidence-Based Policy

In the first decade of this century, there was widespread acknowledgement among analysts and policymakers that interdependencies among problems in related policy fields—energy, environment, public health, social welfare, national security, criminal justice, and economic development—appear to be increasing at an increasing rate, much as a river accumulates force and significance as it traverses falling cascades. In the area of health policy, which is one of the most rapidly growing sectors in the United States, the number of organizations devoted to health policy research and policy advocacy has grown at a rate of approximately 17 percent annually since the turn of this century. It is noteworthy, and perhaps prophetic, that this is the approximate rate of growth of health expenditures as a percentage of GDP in the past 20 years. By the year 2020, the number of these organizations may surpass 15,000.

One response to this and other instances of accelerating interdependency is the development of different systemic visions of these dynamic interdependencies. The 2010 National

---

[31]Twenty-five years ago, Michael Marien, editor of *Future Survey* (the journal of record of the World Future Society), estimated the number of policy journals to exceed 400. Michael Marien, in "The Scope of Policy Studies: Reclaiming Lasswell's Lost Vision," in *Advances in Policy Studies since 1950,* vol. 10, *Policy Studies Review Annual*, ed. William N. Dunn and Rita Mae Kelly (New Brunswick, NJ: Transaction, 1991), pp. 445–488.

[32]The Gale Group web page (http://galegroup.com/) states that *The Encyclopedia of Associations* (*EA*) is the only comprehensive source of detailed knowledge concerning nonprofit American research, professional, and membership organizations of national scope. *EA* is a guide to more than 23,000 organizations in subjects such as Trade, Business, Commercial; Legal, Governmental, Public Administration, Military; Cultural; Educational; Health and Medicine, Veterans, Sports, and many more.

Academy of Medicine report characterizes the world as a complex system composed of a large number of influences that, for the most part, are interacting simultaneously. A systems perspective encourages policymakers and analysts to think expansively, rather than focus on component parts in isolation. A systems perspective can enhance the ability to develop and use evidence effectively and suggest actions with the potential to effect change. A systems perspective "can forecast the potential consequences of not taking action, possible unintended effects of interventions, the likely magnitude of the effect of one or more interventions, conflicts between or complementarity of interventions, and priorities among interventions."[33]

The growth of systems thinking has paralleled the call for *evidence-based policy,* first in the United Kingdom and then in the United States and the European Union. Evidence-based policy is a response to complexity, but at a meso- or micro-level of analysis. On one hand, there is an acknowledgement that ideological, religious, and political influences, often hidden and lacking in transparency, have exerted a potentially harmful effect on our understanding of the factors responsible for unsuccessful policies. In the words of a British House of Commons report called *Scientific Advice, Risk and Evidence Based Policy Making* (2006), evidence-based policymaking "has its roots in Government's commitment to 'what works' over ideologically driven policy. . . . This Government expects . . . new ideas, more willingness to question inherited ways of doing things, better use of evidence and research in policymaking."[34]

Evidence-based policymaking, although it takes different forms, is united in its definition of evidence as causally relevant knowledge of relations between policies, on one hand, and short-term, intermediate, and long-term policy outcomes on the other. Among the largest and most competent repositories of evidence-based policies in health policy and medicine are the Cochrane Collaboration (C1) (www.cochranecollaboration.org) and the Campbell Collaboration (C2) (www.campbellcollaboration.org). C1 and C2 combine a concern with assessing the causally relevant outcomes of policies (and also medical interventions in the case of C1) and the analysis of extant world research on outcomes of these interventions. The two main methodologies used by policy researchers working with C1 and C2 are research synthesis and meta-analysis,[35] which are exceptionally powerful and appropriate methods for analyzing the global stock of knowledge on a subject.

In the United States, evidence-based policy has been promoted by leading program evaluators and policy analysts associated with the Coalition for Evidence-Based Policy of the Council for Excellence in Government.[36] Some procedures of the Office of Management and Budget are based on evidence-based policy analysis, especially methods and standards of program evaluation. In addition, a brief review of recent titles of books on public policy,

[33]Quoted in *Using Science as Evidence in Public Policy*, Committee on the Use of Social Science Knowledge in Public Policy, Division of Behavioral and Social Sciences and Education (Washington, DC: The National Academies Press), p. 60.

[34]United Kingdom, House of Commons, Science and Technology Committee, *Scientific Advice, Risk and Evidence Based Policy Making.* Seventh Report of Session 2005–06, Volume I (London: HMO Printing House, 2006).

[35]A comprehensive text on these two methodologies is Harris M. Cooper, *Research Synthesis and Meta-Analysis: A Step-by-Step Approach*, 4th ed. (Thousand Oaks, CA: Sage Publications, 2010).

[36]Coalition for Evidence Based Policy, www.evidencebasedprograms.org/.

social policy, education policy, security policy, and so forth, shows that the titles begin with the phrase "evidence-based." In this context, it should also be said that some see the movement toward evidence-based policy as a new and harmful logical positivist ("scientistic") approach to questions of public policy, an approach that may be inconsistent with democracy.[37] Others see evidence-based policy as only one of several alternatives to conducting research on public policy, including approaches grounded in qualitative, hermeneutic, and interpretive inquiry.[38]

THE POLICYMAKING PROCESS

As we have seen, the development of policy analysis has been a response to practical problems. Although recognition of its practical origins is essential for understanding policy analysis, historical awareness alone does not tell us much about the characteristics of policymaking and how it works. For this we must turn to the founders of the movement known as the *policy sciences*.

The policy sciences movement was founded by the preeminent political scientist, Harold D. Lasswell, who offered a new vision of the relation between the social sciences and policymaking. Although fragments of Lasswell's vision may be traced to the nineteenth century, the policy sciences did not only mandate the creation of knowledge about policymaking, something that had been done for several centuries. The policy sciences also required that the knowledge that was created be used to improve the process of policymaking.[39] Like Robert S. Lynd's *Knowledge for What*? (1939), this was a revolutionary vision, one that affirmed the centrality of the social sciences in the process of policymaking.

For the policy sciences, analysis was seen as an *intellectual* activity embedded in a *social* context. For Lasswell and Myers S. McDougall, his decades-long collaborator in the development of the policy sciences, this process was divided into three components. The first component is *contextuality*, which refers to the fact that decisions are embedded in a wider social process. The second is *problem orientation*. Practitioners of the policy sciences are proficient in the intellectual activities involved in clarifying goals, trends, conditions, projections, and alternatives. The third is *methodological diversity*, which stresses that the methods employed are broad rather than narrow in range (Lasswell 1971: 4).

The founders of the policy sciences represented the decision process as seven sequentially ordered *decisional functions*, each based on the coercive and persuasive powers of government. Each decisional function has outcomes, and each outcome has effects. Values

[37]Wayne Parsons, "From Muddling Through to Muddling Up: Evidence Based Policy-Making and the Modernisation of British Government," *Public Policy and Administration* 17, 3 (2002): 43–60.
[38]John W. Creswell, *Qualitative Inquiry and Research Design: Choosing Among Five Approaches* (Thousand Oaks, CA: Sage Publications, 2013).
[39]The policy sciences, stated Lasswell (*A Pre-view of Policy Sciences*, 1971: 1), "are concerned with knowledge *of* and *in* the decision processes of the public and civic order."

identified by Lasswell and Kaplan[40]—for example, enlightenment and skill—are effects of outcomes such as gathering, processing, and disseminating knowledge.[41] Here are the seven sequentially ordered decisional functions:

> *Intelligence*. Outcomes of the intelligence function include gathering, processing, and disseminating knowledge for the use of participants in the decision process, for example regulatory agencies and judicial bodies. Values of enlightenment, wealth, and skill are among the possible effects of gathering, processing, and disseminating knowledge.
>
> *Promotion*. Outcomes of the promotion function include agitation and propaganda for the use of leaders, political parties, and interest groups. Values of power and skill are among the possible effects of agitation and propaganda.
>
> *Prescription*. Outcomes of the prescriptive function include the routinization and institutionalization of enforceable norms by executives and legislators. Values of power, respect, and skill are among the possible effects of executive orders, legislative statutes, and other enforceable norms.
>
> *Invocation*. Outcomes of the invocation function include enforced conformity to prescriptions by staff in line agencies. Among the possible effects of enforced conformity to prescriptions are wealth, health, and well-being.
>
> *Application*. Outcomes of the application function include the characterization, or written documentation, by appellate court judges of conditions under which prescriptions are applied. Values of skill, respect, and rectitude are possible effects of cases written by appellate court judges.
>
> *Termination*. Outcomes of the termination function include the adjudication of claims of parties affected by the cancellation of prescriptions. Values of power and well-being are among the possible effects of adjudicating cases before small claims courts.[42]
>
> *Appraisal*. Outcomes of the appraisal function include the assignment of legal or administrative responsibility for the assessment of policy objectives. Values of wealth, power, and enlightenment are possible effects of the assessment of objectives by means of investigative commissions, inspectors general, and research offices of legislatures.

The seven original decisional functions were later revised in Lasswell's *A Pre-view of Policy Sciences,* published in 1971, which was his last major work on the decision process before

[40]Harold D. Lasswell and Abraham Kaplan, *Power and Society: A Framework for Political Inquiry* (New Haven, CT: Yale University Press, 1951). The ten values (or "base-values") include wealth, enlightenment, power, skill, and well-being.

[41]The distinction between outcomes and effects of outcomes is awkward, as originally stated by Lasswell. Another way to state the distinction is to use the terms outcomes (short-term effects) and impacts (long-term effects), as is often done in public program evaluation.

[42]Critics sometimes note that it is illogical that Lasswell placed termination before appraisal. However, "termination" refers to what we now might call "late-stage-implementation." Appraisal then becomes necessary in order to evaluate the outcomes, for example, outcomes in the form of claims adjudicated in a small claims court.

his death in 1978. At first glance, the book appears to present a typical theory of sequential decision-making. Such theories begin with a policy problem, followed by a sequence of activities directed toward its solution. A widely known reworking of these decisional functions includes seven stages of agenda setting, policy formulation, policy adoption, policy implementation, policy communication, policy evaluation, and policy adaptation.[43]

However, Lasswell's sequenced decisions are called functions, not stages or phases, a difference of pivotal theoretical significance, because Lasswell followed John Dewey in viewing a function as a purposeful (or teleological) act, not simply a set of temporally ordered activities. For Dewey, a function is "any process sufficiently complex to involve an arrangement or coordination of minor processes which *fulfills a specific end* in such a way as to conserve itself."[44] Functions, viewed in this way, may be traced to a tradition of practical reasoning documented by Georg Henrik von Wright in his treatise on epistemological theories associated with logical positivism and interpretivism.[45] Achieving a purpose is a consequence of making decisions, or choices, about future states of affairs, rather than applying inductive or deductive rules in advance.[46] Policymaking, write Lasswell and McDougal, "is not a body of rules, but . . . a process of making authoritative decisions about the production and distribution of values in the community. . . . Rules alone could not explain why past decisions had been made or how future decisions were likely to be made."[47]

The legacy of Lasswell, McDougall, and the policy sciences is what we know now as the *policymaking process,* or *policy process* for short. Today, as in Lasswell's time, it is useful to visualize this process as a series of interdependent functions arrayed through time. The process begins with agenda setting, policy formulation, and policy adoption and proceeds through policy implementation, policy assessment, and policy adaptation (Table 2.1).[48] Depending on circumstances, analysts produce knowledge relevant to one, several, or all phases of policymaking.

The policy process is composed of complex cycles (Figure 2.1). Each phase of a *policy cycle* is linked to the next, in backward and forward loops, and the process as whole has no definite beginning or end. Individuals, interest groups, bureaus, departments, and

---

[43] Although they differ slightly, the first reworking of the decisional functions was Charles O. Jones, *An Introduction to the Study of Public Policy*, 2nd ed. (North Scituate, MA: Duxbury Press, 1977), followed by James A. Anderson, *Public Policymaking: An Introduction*, 7th ed. (Boston, MA: Wadsworth, 2011). Each are organized in accordance with similar stages of the policymaking process that are based on Lasswell.

[44] John Dewey, "Brief Studies in Realism. II," *The Journal of Philosophy, Psychology and Scientific Methods* 8, 20 (1911): 546–554.

[45] Georg Henrik von Wright, *Explanation and Understanding* (Ithaca, NY: Cornell University Press, 1971).

[46] McDougal, Myres S., Harold D. Lasswell, and W. Michael Reisman, "The Intelligence Function and World Public Order." *Temple Law Quarterly* 46 (1972): 365.

[47] Lasswell, Harold D., and Myres S. McDougal, "Legal Education and Public Policy: Professional Training in the Public Interest," *The Yale Law Journal* 52, 2 (1943): 203–295.

[48] The policy cycle, or stages approach, is summarized by Peter de Leon, "The Stages Approach to the Policy Process: What Has It Done? Where Is It Going?" in *Theories of the Policy Process*, ed. Paul A. Sabatier (Boulder, CO: Westview Press, 1999). See also Charles O. Jones, *An Introduction to the Study of Public Policy*, 2nd ed. (North Scituate, MA: Duxbury Press, 1977); James A. Anderson, *Public Policymaking* (New York: Praeger, 1975); and Gary Brewer and Peter de Leon, *Foundations of Policy Analysis* (Homewood, IL: Dorsey Press, 1983). The classic work that influenced all works cited here is Harold D. Lasswell, *The Decision Process: Seven Categories of Functional Analysis* (College Park, MD: Bureau of Governmental Research, University of Maryland, 1956).

TABLE 2.1

Phases of the Policymaking Process

Phase	Characteristics	Illustration
Agenda Setting	Under pressure from organized interests, elected and appointed officials place problems on the governmental agenda. Many problems are not acted on at all, while others are addressed after long periods.	A state legislator, at the request of a local welfare rights organization, passes to staff analysts a legislative inquiry regarding the issue of a living wage. Staff analysts brief members of the Health and Welfare Committee, who show no interest in legislation. The issue is not acted on by government for twenty years.
Policy Formulation	Officials formulate alternative policies to deal with a problem. Alternative policies assume the form of executive orders, court decisions, and legislative acts.	A state court considers prohibiting the use of standardized achievement tests such as the Scholastic Aptitude Test (SAT) on grounds that the tests are biased against women and minorities.
Policy Adoption	A policy is adopted with the support of a legislative majority, consensus among age agency directors, or a majority court decision.	In *Roe v. Wade*, Supreme Court justices reach a majority decision that women have the right to terminate pregnancies through abortion.
Policy Implementation	An adopted policy is carried out by administrative units that mobilize financial and human resources to implement the policy.	The city treasurer hires additional staff to ensure compliance with a new law that imposes taxes on hospitals that no longer have tax-exempt status.
Policy Assessment	Auditing and accounting units in government determine whether executive orders, legislative acts, and court decisions are in compliance with statutory requirements and realizing their objectives.	The U.S. General Accountability Office monitors social welfare programs such as Temporary Assistance for Needy Families (TANF) to determine the scope of potential welfare fraud.
Policy Adaptation	Auditing and evaluation units report to agencies responsible for formulating, adopting, and implementing policies that poorly written regulations, insufficient resources, and inadequate training require the adaptation of policies.	A state Department of Labor and Industry evaluates an affirmative action training program, finding that employees wrongly believe that complaints against discrimination should be made to immediate supervisors, who are the objects of complaints, rather than to affirmative action officers.
Policy Succession	Agencies responsible for evaluating policies determine that a policy is no longer needed because the problem has been resolved. Rather than terminate the policy, its objectives are redefined.	The National Highway Traffic Safety Administration (NHTSA) convinces Congress to maintain the 55 mph speed limit, originally designed to save fuel, because it is unexpectedly reducing traffic fatalities.
Policy Termination	A legislative committee responsible for evaluation and oversight of agencies determines that a policy or an entire agency should be terminated because it is no longer needed.	The U.S. Congress terminates the Office of Technology Assessment (OTA) on grounds that it is not needed because other agencies and the private sector are able to assess the economic and social effects of technologies.

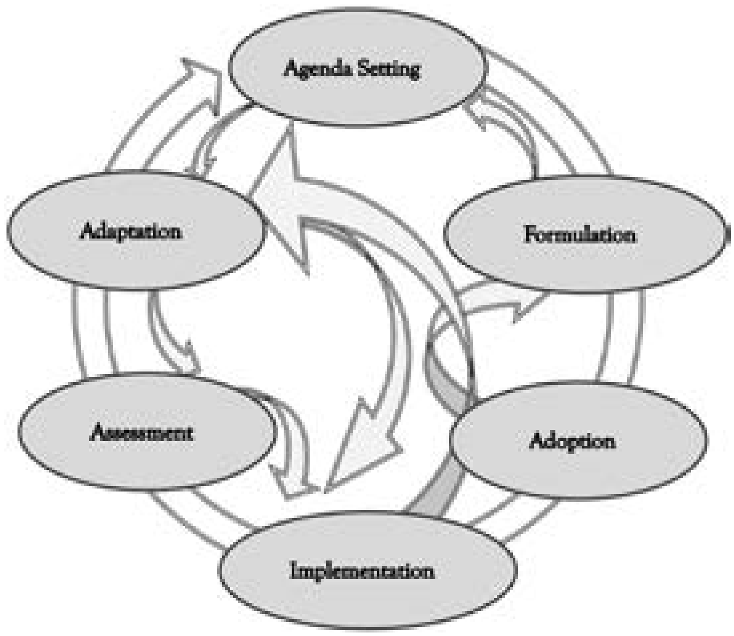


FIGURE 2.1
Complexity, Feedback, and Short Circuiting in the Policymaking Process

ministries are involved in policy cycles through cooperation, competition, and conflict. One form of cycle involves *policy adaptation,* where a feedback loop connects to earlier phases. Other forms of cycles are *policy succession,* where new policies and organizations build on old ones, and *policy termination.* Policy termination may mean the end of a policy or program, although even termination affects what issues are placed on the public agenda and in this sense represents another kind of cycle.

In some cases, a policy is adopted first, and then justified by working backward to agenda setting, where a problem is formulated or reformulated to fit or justify the policy. Parallel cycles occur, where different groups develop policies at the same time, and there may be forward ("arborescent") as well as backward ("assembly") branching from one phase to multiple successor or predecessor phases. Adjacent phases may be linked, or skipped altogether, creating "short circuits." Solutions and problems are in continuous flux, creating "primeval policy soups" and "organized anarchies."[49] Figure 2.1 also assumes that policy-relevant knowledge may affect the policymaking process, although there is no assurance that this will occur.

[49]Michael Cohen, James March, and Johan Olsen, "A Garbage Can Model of Organizational Choice," *Administrative Science Quarterly* 17 (March 1972): 1–25; John W. Kingdon, *Agendas, Alternatives, and Public Policies*, 2nd ed. (New York: Harper Collins, 1995).

MODELS OF POLICY CHANGE

Metaphors of garbage cans, primeval policy soups, and organized anarchies are easily visualized, but they are difficult to apply to concrete policies. Indeed, it may seem that policies, when represented in this way, have no structure or form at all. However, a story may help to clarify the importance of these metaphors.[50]

> One month ago, on a secret military base, there were reports of the theft of classified materials. Every day, at about the same time, a worker pushing a wheelbarrow would attempt to leave the base by passing by the security guard. On the first day, the guard asked the worker what was in the wheelbarrow. He replied, "Just dirt." The guard probed the dirt with his baton and, finding nothing, allowed the worker to pass. The probing was repeated successfully for 1 week. During the second week, however, the guards grew suspicious, ordering the worker to dump the contents of the wheelbarrow on the ground. This time, the guards used a rake to examine the contents, again finding nothing. This continued for another 2 weeks.
>
> A week later the guards summoned a special investigation unit equipped with a scanning device. The worker was again ordered to dump the contents, which were scanned, but with no results. Thereafter, the worker was allowed to pass freely. At the end of the month, thirty wheelbarrows were reported missing.

The point of the story is that structures are just as important as the processes they are intended to represent. Conceptual models, which are particular kinds of intellectual structures, are essential for understanding the process of policymaking. These conceptual models, which are frequently based on metaphors such as garbage cans, organized anarchies, and primeval soups,[51] structure the ways we think about and explain processes of policymaking.

The Comprehensive Rationality Model

The *comprehensive rationality* model structures the way we think about and explain the process of decision-making. This model explains policymaking as an exhaustive striving for economic efficiency. A rational economic actor, or *homo economicus,* weighs the costs and benefits of available alternatives and makes a decision that is motivated by a concern with the efficient use of resources. The basic proposition of the comprehensive rationality model is as follows: Among two or more exhaustive alternatives, the alternative selected will (or should) have net efficiency benefits (benefits less costs) that exceed the other alternatives.[52] Economically rational decisions are made when decision-makers:

- Specify an exhaustive set of alternatives on which there is sufficient consensus to act on behalf of others.

[50]I owe this example to John Funari, personal communication.

[51]Other metaphors and similes are "the wall," "surge," "policy decay," "war on poverty," and "problems as infectious diseases." Rein and Schon call them "generative metaphors" because they generate a framework for structuring policy problems. See Martin Rein and Donald A. Schon, "Problem Setting in Policy Research," in *Using Social Research in Public Policymaking*, ed. Carol Weiss (Lexington, MA: D. C. Heath, 1977), pp. 240–243.

[52]Rational choice models are reviewed by Elinor Ostrom, "Institutional Rational Choice: An Assessment of the Institutional Analysis and Development Framework," in *Theories of the Policy Process,* ed. Sabatier, pp.

- Identify objectives for each alternative.
- Forecast the consequences of selecting each alternative.
- Specify and rank transitively alternatives that best achieve objectives.
- Choose the alternative that maximizes the attainment of objectives.

There are other versions of rational choice that involve the incorporation of institutional transaction costs into choice situations[53]; the redefinition of benefits and costs in political, social, or ethical terms (e.g., *homo politicus*)[54]; and the qualification that decision-makers are fallible learners with imperfect knowledge, limited computational capabilities, and a propensity for error.[55]

Second-Best Rationality

An important criticism of the model of comprehensive economic rationality is *Arrow's impossibility theorem,* which provides an alternative explanation of policymaking processes. The theorem states that it is impossible for decision-makers in a democratic society to meet the requirements of comprehensive economic rationality.[56] It is impossible to make a single best choice by aggregating individual rational choices by means of majority voting. The impossibility of creating a collective decision that involves transitive preferences (if A is preferred to B, and B is preferred to C, then A is preferred to C) is described as the "voters' paradox."

Consider a committee composed of three members: Brown, Jones, and Smith. The committee wants to decide which of three forms of energy—solar, coal, and nuclear—should be adopted to resolve the energy crisis. Brown, the leader of an energy rights organization, prefers solar to coal and coal to nuclear, reasoning that this ranking creates least risk to citizens. Because Brown's choice is transitive, it conforms to the rule: If A is preferred to B, and B is preferred to C, then A is preferred to C. Jones and Smith, who represent the coal and nuclear industries, also have transitive preferences. Jones prefers coal to nuclear, nuclear to solar, and coal to solar, reasoning that coal is most profitable, followed by nuclear and solar. In turn, Smith prefers nuclear to solar, solar to coal, and nuclear to coal, reasoning that nuclear is most profitable. Solar energy is less profitable than coal but creates fewer dangers to the environment. Smith sees coal as the least desirable of the three alternatives.

The three choices are rational and transitive from each individual's point of view. Yet, once the three members attempt through majority rule to reach a democratic decision, there is a paradox (Table 2.2). When we ask them to choose between solar and coal, we see that solar is preferred to coal by a margin of 2 to 1 (Brown and Smith versus Jones). Similarly, when we ask them to choose between coal and nuclear, we observe that coal is

35–72. The classic critique of comprehensive rational choice is Charles E. Lindblom, *The Policy-Making Process* (Englewood Cliffs, NJ: Prentice Hall, 1968).

[53]Oliver E. Williamson, *The Economic Institutions of Capitalism* (New York: Free Press, 1985).

[54]David J. Silverman, *The Theory of Organisations* (New York: Free Press, 1972).

[55]Elinor Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action* (New York: Cambridge University Press, 1990), and Mancur Olson, *The Logic of Collective Action: Public Goods and the Theory of Groups* (Cambridge, MA: Harvard University Press, 1965).

[56]Kenneth J. Arrow, *Social Choice and Individual Values* (New York: John Wiley, 1963).

TABLE 2.2

The Voters' Paradox

Committee Member	Preference
Brown	A (solar) preferred to B (coal)
	B (coal) preferred to C (nuclear)
	A (solar) preferred to C (nuclear)
Jones	B (coal) preferred to C (nuclear)
	C (nuclear) preferred to A (solar)
	B (coal) preferred to A (solar)
Smith	C (nuclear) preferred to A (solar)
	A (solar) preferred to B (coal)
	C (nuclear) preferred to B (coal)
Majority	A (solar) preferred to B (coal)
	B (coal) preferred to C (nuclear)
	C (nuclear) preferred to A (solar)

preferred to nuclear by a 2 to 1 margin (Brown and Jones versus Smith). However, if we now apply the transitivity rule to these *collective* preferences, we should find that if A is preferred to B (solar to coal), and B is preferred to C (coal to nuclear), then A is preferred to C (solar to nuclear). However, this is not the collective result, because two committee members (Jones and Smith) prefer C to A (nuclear to solar). Hence, individual preferences are transitive, while the collective preference is *cyclic*, which means that alternatives cannot be ranked consistently. For this reason, a rational choice is impossible.

Arrow's impossibility theorem proves by logical deduction that it is impossible to apply democratic procedures (e.g., majority rule) to reach collective decisions that are transitive. Specifically, it is impossible to meet one or more "reasonable conditions" of any democratic decision procedure: (1) *nonrestriction of choices*, that is, all possible combinations of individual preferences must be taken into account; (2) *nonperversity of collective choice,* that is, collective choices must consistently reflect individual choices; (3) *independence of irrelevant alternatives*, that is, choices must be confined to a given set of alternatives that are independent of all others; (4) *citizens' sovereignty*, that is, collective choices must not be constrained by prior choices; and (5) *nondictatorship*, that is, no individual or group can determine the outcome of collective choices by imposing its preferences on others.

To avoid the dilemma of intransitive preferences, collective choices might be delegated to a few decision-makers (e.g., political or technical elites) who can be expected to reach consensus and make a transitive choice. Although this solves the problem of intransitive preferences, it nevertheless violates conditions of citizens' sovereignty and nondictatorship. Alternatively, we might introduce additional alternatives (e.g., wind power) in the hope that this will foster consensus, except this violates the condition of independence of irrelevant alternatives. In practice, however, political systems based on majority rule employ both procedures (delegation and adding alternatives) to reach collective choices.[57] These choices are known as *second best* decisions.

Disjointed Incrementalism

The *disjointed-incremental model* holds that the comprehensive economic rationality model is unworkable as an explanation of policymaking processes.[58] The fundamental proposition of disjointed-incremental theory is that decisions are made at the margin of the status quo, so that behavior at time $t + 1$ is marginally different from behavior at time t. According to disjointed incrementalism, rational choices are made when decision-makers:

- Consider only those alternatives that differ incrementally (i.e., by small amounts) from the status quo.
- Limit the number of consequences forecast for each alternative.
- Mutually adjust objectives, on one hand, and policy alternatives on the other.
- Reformulate problems and alternatives in the course of acquiring new knowledge.
- Analyze and evaluate alternatives sequentially, so that choices are amended over time, rather than at a single point in time.
- Remedy existing problems, rather than attempt to solve problems completely.
- Share responsibilities for analysis and evaluation with multiple groups, creating a process that is disjointed and not conjointed.

Bounded Rationality

Closely related to disjointed incrementalism is the *bounded rationality* model, which offers an alternative explanation of policymaking processes. However, whereas disjointed incrementalism is about the limitations on collective rational choices, bounded rationality is about the limitations on individual rational choices. Bounded rationality asserts that individual decision-makers do not attempt to be rational in the full, or comprehensive, sense of considering and weighing all alternatives.[59] Although choices are rational, they are bounded by the practical circumstances and constraints of complex decisions. The basic proposition of bounded rationality is that decisions are based on "rules of thumb" for making minimally acceptable choices. The originator of the bounded rationality model, Herbert Simon, argues that it is impossible for a single, isolated individual to reach any high degree of rationality. "The number of alternatives he must explore is so great, the knowledge he would need

---

[57]The setting of the sequence of issues on agendas is an important example of violating conditions of citizens' sovereignty and nondictatorship. See Duncan Black, *The Theory of Committees and Elections* (Cambridge: Cambridge University Press, 1958). For a review of these problems, see Norman Frohlich and Joe A. Oppenheimer, *Modern Political Economy* (Englewood Cliffs, NJ: Prentice Hall, 1978), Ch. 1.

[58]Charles E. Lindblom and David Braybrooke, *A Strategy of Decision* (New York: Free Press, 1963).

[59]See Herbert A. Simon, *Administrative Behavior* (New York: Macmillan, 1945). In 1978, Simon received the Nobel Prize in Economics for his contributions to the study of decision-making in economic organizations. Some of Simon's other important works are *Models of Man* (New York: Wiley, 1957) and *The Sciences of the Artificial* (New York: Wiley, 1970).

to evaluate them so vast, that even an approximation to objective rationality is hard to conceive."[60]

In contrast to *maximizing behavior,* Simon proposes the concept of *satisficing behavior*. Satisficing refers to decisions that are just "good enough," that is, where decisions combine *satisfactory* and *suffice* to create a *satisficing* choice. In other words, decision-makers do not consider the full range of alternatives that might maximize the achievement of objectives. To make *satisficing* decisions, policymakers and analysts consider only the most readily identifiable potential solutions.

Some see satisficing behavior as an effort to maximize valued outcomes while recognizing the constraints imposed by the costs of information. Considering that the search for information is time-consuming and costly, "an optimal decision procedure should take these factors into account . . . the cost of decision-making should be incorporated into the maximizing model."[61]

Erotetic Rationality

In simplest terms, erotetic rationality refers to a process of questioning and answering.[62] Models of erotetic rationality, although they may be unsatisfying to those who demand answers from deductive and inductive-statistical models of decision processes, originate in an entirely different methodological tradition, that of the pragmatism of James, Peirce, and Dewey in philosophy and that of Harold Lasswell, Myres McDougal, and Donald Schon in public policy and the policy sciences.[63] All claim that much of real-world policymaking conforms to processes of abductive reasoning, not those of deduction and induction.[64]

An important question about erotetic rationality is where the process of questioning begins. One answer may be found in the history of methodological pragmatism, especially as that history is found in the works of Charles Sanders Peirce, for whom questions arise in the process of abduction. As Bromley emphasizes, abduction kicks in when we are faced with a surprise and, more generally, when our beliefs are "unsettled." Another setting in which beliefs are unsettled is in what Dewey called a *problem situation.* Rein and Schon, drawing on

[60]Simon, *Administrative Behavior*, p. 79.

[61]Richard Zeckhauser and Elmer Schaefer, "Public Policy and Normative Economic Theory," in *The Study of Policy Formation,* ed. Raymond A. Bauer (Glencoe, IL: Free Press, 1966), p. 92.

[62]Nicholas Rescher, a methodological pragmatist, makes a case for erotetic rationality in *Induction* (Pittsburgh, PA: University of Pittsburgh Press, 1980), pp. 6–7.

[63]The description of pragmatism follows Nicholas Rescher, *Methodological Pragmatism: A Systems-Theoretic Approach to the Theory of Knowledge* (Oxford: Basil Blackwell 1971), who prefers the term *methodological pragmatism*, a term that distinguishes it from the vulgar notion that pragmatism means "practical," and from the more abstract emphases on philosophical aspects including ontology and epistemology. Models of erotetic rationality, although they may be unsatisfying to those who demand well-specified deductive and statistical models of decision processes, claim that much real-world policymaking conforms to processes of abductive reasoning and action.

[64]A persuasive case for this claim has been made by Daniel W. Bromley, an environmental and institutional economist, who grounds his pragmatist model of policymaking in the practice of abductive reasoning. See Daniel W. Bromley, *Sufficient Reason: Volitional Pragmatism and the Meaning of Economic Institutions* (Princeton, NJ: Princeton University Press, 2006).

Dewey's distinction between a problem and a problem situation, describe a problem situation (not a problem) as "a diffuse set of worries and inchoate signs of stress."[65] A similar formulation of unsettled belief is that of Albert, who contends that ignorance is the sine qua non of rationality in conducting cost-benefit analyses.[66] Here as elsewhere, the beliefs of analysts and policymakers are "unsettled" when they acknowledge that the relationship between policies, policy outcomes, and social values are unknown or indeterminate. Here, the frank acknowledgement of ignorance is a prerequisite of engaging in a process of questioning and answering, a process that yields rationally optimal answers to questions that "transcend accreted experience and outrun the reach of the knowledge already at our disposal."[67]

Bromley provides a compelling example of the process of abduction, which he links to inferences based on the practical syllogism, which like all syllogisms has three elements or premises. Consider the example of a person turning on a light switch to resolve his unsettled belief in the properties of objects that are insufficiently illuminated. The person presses the light switch because he wants the room to be illuminated. This desire for a future state (illumination) is the reason for the choice, while the cause is that he pressed the light switch. The pressing of the switch "is merely a necessary though quite uninteresting step in the process that starts with reason, entails a causal sequence, and ends with a desired outcome" (Bromley 2006: 7).

The process of questioning and answering, or erotetic rationality, is closely related to problem structuring as the central guidance system of policy analysis (see Chapter 3). Erotetic rationality lies at the core of recent innovations in physics, as represented by the work of Ilya Prigogine, the originator of chaos theory: Physics is no longer about certainties, but of questioning-and-answering, "everywhere, instead of uncovering the permanence and immutability that classical science has taught us to seek out in nature, we have encountered change, instability, and evolution."[68] Students of policy who interrogate practice, abandoning deductive and inductive-statistical models,[69] appear to be describing the process of erotetic reasoning and abduction.[70]

Simultaneous Convergence

Simultaneous convergence refers to processes that are like a complex river delta with multiple streams converging and diverging as they cross the flood plain toward the sea.[71] If we try to understand the "outcomes" of the process, by using a bucket of water to sample from

[65]Martin Rein and Donald Schon, "Problem Setting in Policy Research," *Using Social Research in Public Policymaking* (1977): 235–236.
[66]Jeffrey M. Albert, "Some Epistemological Aspects of Cost-Benefit Analysis," *George Washington Law Review* 45, 5 (1977): 1030.
[67]Rescher, *Induction,* p. 6.
[68]Ilya Prigogine, "A New Model of Time, a New View of Physics," in *Models of Reality,* ed. Jacques Richardson (Mt. Airy, MD: Lomond Publications, 1984). Quoted by Rita Mae Kelly and William N. Dunn, "Some Final Thoughts," in *Advances in Policy Studies since 1950,* ed. Dunn and Kelly, p. 526.
[69]Hal K. Colebatch, Robert Hoppe, and Mirko Noordegraaf, *Working for Policy* (Amsterdam: Amsterdam University Press, 2010).
[70]Another definition of abduction is "the method of hypothesis" (Peirce).
[71]Alex Weilenmann attributes the metaphor to systems theorist Stafford Beer. Personal communication.

the sea, we will not be able to discover the various routes, or structures, by which the water reaches the sea. However, if we were able to monitor the entire process and its structure, we would find that from time to time some streams converge to form deep channels that, at least temporarily, guide the process in predictable ways.

The metaphor of the river delta and its multiple streams conveys some of the complexity of the policymaking process, but without abandoning the responsibility to identify structures that shape the process. In policymaking contexts, individuals and groups interact over time to set agendas and formulate policies. Their success, however, depends on the ability to recognize critical moments ("policy windows") when three kinds of streams—problems, policies, and politics—converge.[72] The fundamental proposition of the critical convergence model is that *policy change occurs at these critical moments.* The recognition of these moments is part of the challenge facing the analyst.

Punctuated Equilibrium

One of the difficulties of many of these models is that none satisfactorily accounts for major departures from the dominant pattern of gradual change predicted by the disjointed-incremental model. One major departure was the abrupt, discontinuous, and relatively rare changes that occurred in environmental policy under the Clean Air Act.[73] Recognizing that significant policy change may occur only every 25 years or so, Sabatier and Jenkins-Smith[74] call attention to the importance of external perturbations, or exogenous shocks, in effecting policy changes that are discontinuous and broad in scope. Such changes include relatively rapid socioeconomic changes represented by recessions, depressions, and oil crises; sudden shifts in public opinion such as those that occurred late in the Vietnam War; and the dramatic rise in nationalism and personal insecurity that occurred after the September 11, 2001, attacks on the World Trade Center and the Pentagon.

The *punctuated equilibrium model* likens the process of policy change to biological evolution.[75] Most policies are relatively stable, changing incrementally over long periods. There is a dynamic equilibrium among competing policies, much like the process of partisan mutual adjustment identified by Lindblom and Braybrooke.[76] Partisan mutual adjustment can and often does involve a competitive process among relatively autonomous bureaucracies. Bureaucratic leaders compete for recognition and resources in accordance with the

[72]In contrast to John Kingdon, *Agendas, Alternatives, and Public Policies*, I am using "river delta" in place of "garbage can." The combination of "streams," "garbage cans," and "windows" is a seriously mixed metaphor that I wish to avoid.

[73]See Charles O. Jones, *Clean Air* (Pittsburgh, PA: University of Pittsburgh Press, 1975).

[74]Paul A. Sabatier and Hank C. Jenkins-Smith, "The Advocacy Coalition Framework: An Assessment," in *Theories of the Policy Process,* ed. Sabatier, pp. 117–166. Two causally relevant conditions affecting policy change are external shocks and core values. Their "advocacy coalition framework" parallels theories of population dynamics in biology.

[75]See James L. True, Frank R. Baumgartner, and Bryan D. Jones, "Punctuated-Equilibrium Theory: Explaining Stability and Change in American Policymaking," in *Theories of the Policy Process,* ed. Sabatier, pp. 97–116, and Frank R. Baumgartner and Bryan D. Jones, *Agendas and Instability in American Politics* (Chicago, IL: University of Chicago Press, 1993).

[76]Charles E. Lindblom and David Braybrooke, *A Strategy of Decision* (New York: Free Press, 1963).

institutional incentives and reward structure of their organizations. Periodically, abrupt changes (punctuations) in policy occur as a consequence of new political images, which, in turn, are products of "policy earthquakes" and other external shocks. The fundamental proposition of the punctuated equilibrium model is that external shocks are a necessary but not sufficient condition of major policy change. The sufficient condition is that new political images and understandings of the political world arise in response to these shocks. However, when new political images, beliefs, and values develop gradually, over long periods of time, the process is not "punctuated."[77]

POLICY ANALYSIS IN THE POLICYMAKING PROCESS

The purpose of policy analysis is to improve policymaking. This is no simple task, considering that many of the most important policies are resistant even to incremental changes. Large, discontinuous changes are relatively rare, and they usually stem from external shocks to the policy system, not from internal changes prescribed by analysts and acted on by policymakers. Nevertheless, multidisciplinary policy analysis is important because of the potential for change embodied in the five types of policy-relevant knowledge and associated methods (Figure 2.2).

Potential Uses of Analysis

Among many analysts, it is an unspoken article of faith that producing policy-relevant knowledge will yield better policies. Here, the presumed practical efficacy of knowledge is evident in the functions, or purposes, served by the methods.

Problem Structuring. The function of problem-structuring methods is to supply policy-relevant knowledge to those who wish to examine the assumptions underlying the definition of problems on the *public agenda* (see Table 2.1). In the agenda-setting phase of policymaking, problem structuring can assist in discovering hidden assumptions, diagnosing rival causes, mapping possible objectives, synthesizing conflicting views, and visualizing new and better policies. For example, the problem of race and sex bias in some 20 million standardized tests administered annually in the United States was placed on the legislative agendas of several states throughout the late 1980s. In Pennsylvania, analysts challenged the assumption that test bias is a problem requiring legislative action—for example, the outright prohibition of standardized tests—after synthesizing and evaluating research on test bias published in professional and academic literature. There were large observed discrepancies in minority and white achievement test scores, but it was confirmed that most of these differences did not arise from test bias, per se. Rather they were an indicator, or gauge, of inequality of educational opportunity. Analysts recommended the continued use of standardized tests to monitor these inequalities, leading to the removal of the issue from the legislative agenda.[78]

---

[77]See Sabatier and Jenkins-Smith, "The Advocacy Coalition Framework."

[78]William N. Dunn and Gary Roberts, *The Role of Standardized Tests in Minority-Oriented Curricular Reform,* policy paper prepared for the Legislative Office for Research Liaison, Pennsylvania House of Representatives, February 1987.

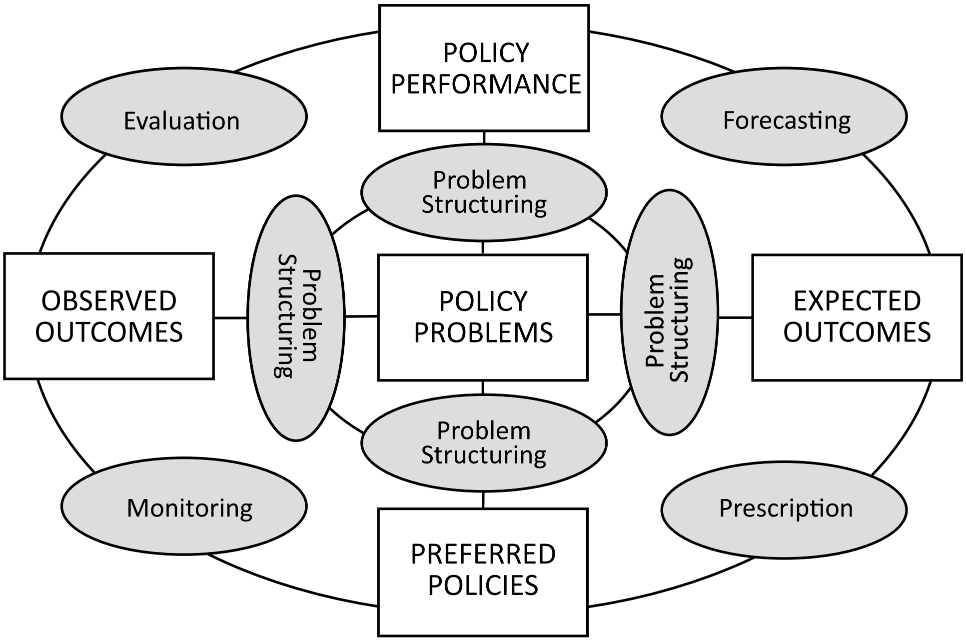


FIGURE 2.2
Policy-Relevant Knowledge and Associated Methods

Forecasting. Methods for forecasting expected policy outcomes provide policy-relevant knowledge about consequences that follow the adoption of preferred policies (including doing nothing) at the phase of *policy formulation*. Forecasting helps examine plausible, potential, and normatively valued futures; estimate the consequences of existing and proposed policies; specify probable future constraints on the achievement of objectives; and estimate the political feasibility (support and opposition) of different options. In 1990, for example, analysts in the Health Care Finance Administration employed forecasting methods to estimate the effects of revenue shortfalls on the Medicare trust fund, which could be exhausted in the next two decades. In the absence of new health care policy initiatives, future benefits under Medicare might be reduced by $40–$50 billion, creating a program reduction of 50 percent. In the meantime, given the baby-boom, the some 40 million persons who had no health insurance were likely to increase in number.[79] The passage of the Affordable Care Act (Abamperes008) significantly reduced the number of persons without health care.

Prescription. Methods for prescribing preferred policy alternatives yield policy-relevant knowledge about the benefits and costs—and more generally the value or utility—of expected policy outcomes estimated through forecasting. This aids policymakers in the

[79]Sally T. Sonnefeld, Daniel R. Waldo, Jeffrey A. Lemieux, and David R. McKusick, "Projections of National Health Expenditures through the Year 2000," *Health Care Financing Review* 13, 1 (Fall 1991): 1–27.

policy adoption phase. By using methods for prescribing preferred policies, analysts may estimate levels of risk and uncertainty, identify externalities and spillovers, specify criteria for making choices, and assign administrative responsibility for implementing policies. For example, the debate over maximum speed limits in the United States has focused on the costs per fatality averted under the 55 and 65 mph speed limits. In the 1980s, one recommendation, based on the belief that the 55 mph speed limit will continue to account for no more than 2–3 percent of fatalities averted, prescribed a shift in expenditures for maintaining the speed limit to the purchase of smoke detectors, which would save many more lives.[80] By 1987, some forty states were experimenting with higher speed limits. In 1995, the 55 mph speed limit was abandoned altogether.

Monitoring. Methods for monitoring observed policy outcomes provide knowledge about the consequences of adopting policies, thus assisting in the *policy implementation* phase. Many agencies regularly monitor policy outcomes and impacts by employing policy indicators in areas of health, education, housing, welfare, crime, and science and technology.[81] Monitoring helps assess degrees of compliance, discover unintended consequences of policies and programs, identify implementation obstacles and constraints, and locate sources of responsibility for departures from policies. For example, economic and social welfare policies in the United States are monitored by analysts in several agencies, including the Bureau of the Census and the Bureau of Labor Statistics. A 1991 analysis concluded that real median household income in the United States grew by merely 2 percent between 1969 and 1989. In the same period, the share of national income by the top fifth of households grew from 43 to 46.7 percent, with all other income groups experiencing a decline. In this case, policy monitoring revealed a marked increase in income inequality, an erosion of the middle class, and a decline in the standard of living between 1969 and 1989.[82] The situation in the period 1989–2015 has been marked by greater income inequality.

Evaluation. Methods for evaluating observed policy outcomes yield policy-relevant knowledge about discrepancies between expected and actual policy performance, thus assisting in the *policy assessment, policy adaptation*, *policy succession*, and *policy termination* phases. Evaluation not only results in conclusions about the extent to which problems have been alleviated, but it also may contribute to the clarification and critique of values driving a policy, aid in the adjustment or reformulation of policies, and establish a basis for restructuring problems. A good example of evaluation is the type of analysis that contributes to the

[80]See, for example, Charles A. Lave and Lester B. Lave, "Barriers to Increasing Highway Safety," in *Challenging the Old Order: Towards New Directions in Traffic Safety Theory*, ed. J. Peter Rothe (New Brunswick, NJ: Transaction Books, 1990), pp. 77–94.

[81]See Duncan MacRae Jr., *Policy Indicators: Links between Social Science and Public Debate* (Chapel Hill: University of North Carolina Press, 1985).

[82]Gordon Green, Paul Ryscavage, and Edward Welniak, "Factors Affecting Growing Income Inequality: A Decomposition," paper presented at the 66th Annual Conference of the Western Economic Association International, Seattle, Washington, July 2, 1991.

clarification, critique, and debate of values by proposing that ethical and moral reasoning augment the technical reasoning driving environmental policies in the European Union and other regions and countries.[83]

Uses of Analysis in Practice

The use of policy analysis and its results has the potential for realizing Lasswell's vision of knowledge *of* and *in* the policy process. In practice, however, realizing this vision is subject to limitations.[84]

- *Use is indirect, delayed, and general.* The use of analysis is typically indirect, delayed, and general. Single analyses are rarely viewed as important unless they are part of a larger body of knowledge. The indirect and general character of use is understandable when we recognize that policymaking is a complex process composed of unpredictable cycles, ranging from termination and succession to adaptation and short-circuiting.
- *The meaning of improvement is ethically controversial.* When the results of analysis are used, the meaning of what constitutes an improvement depends on the political, ideological, and ethical perspectives of observers. Those who believe that the public interest is served by assisting those who are worse off by taxing those who are better off may see "improvement" differently from those who believe that the public interest is served when individuals solve problems without government intervention. To be sure, the value of "efficiency" is often seen as something everyone supports. However, efficiency itself is an ethical matter, because policies very often benefit some individuals and impose costs on others.
- *Being useful reflects personal, professional, and institutional interests.* As professionals and as persons, analysts seek to enhance their status, rewards, and well-being, not only those of their agencies and communities. Gaining access to persons with political power, privilege, and economic standing—as their advisers, consultants, expert witnesses, or staff analysts—is part of the motivation for being a professional.

Misunderstandings about the uses of analysis stem from a failure to recognize that the process of using policy-relevant knowledge is just as complex as the process of policymaking itself (Figure 2.1). Consider this description of making a policy decision by Chester Barnard, a successful CEO and major contributor to the theory and practice of public and business administration. Barnard wrote that implementing a policy to move a telephone pole from one side of the street to the other, probably required some 10,000 decisions carried out by

---

[83]See, for example, Silvio O. Funtowicz and Jerome R. Ravetz, "Global Environmental Issues and the Emergence of Second Order Science" (Luxembourg: Commission of the European Communities, Directorate-General for Telecommunications, Knowledge Industries, and Innovation, 1990). See also Funtowicz and Ravetz, "A New Scientific Methodology for Global Environmental Issues," in *Ecological Economics*, ed. Robert Costanza (New York: Columbia University Press, 1991), pp. 137–152.

[84]The following discussion draws on Carol H. Weiss, "Introduction," in *Using Social Research in Public Policymaking,* ed. Weiss, pp. 1–22, and Carol H. Weiss with Michael J. Bucuvalas, *Social Science Research and Decision Making* (New York: Columbia University Press, 1980).

100 employees located in a dozen locations. To perform a competent analysis, it would be necessary to investigate the social, political, moral, legal, economic, and physical facts of the environment, a process "requiring 9,000 redefinitions and refinements of purpose . . . probably not more than half-a-dozen decisions will be recalled or deemed worthy of mention. . . . The others will be 'taken for granted,' all of a part of the business of knowing one's business."[85]

- *Composition of users.* Individuals as well as collectives—for example, agencies, ministries, bureaus, courts, legislatures, parliaments—use policy analysis. When the use of analysis involves gains (or losses) in the personal utility of knowledge, the process of use is an aspect of individual decisions (individual use). By contrast, when the process of use involves many individuals, it is an aspect of collective decisions, or policies (collective use).
- *Scope of knowledge used.* The scope of knowledge used by policymakers ranges from the specific to the general. The use of "ideas in good currency" is general in scope (general use), and includes the use of metaphors and similes that capture the essential meaning of ideas and actions.[86] By contrast, the use of a particular recommendation is specific (specific use), as exemplified by standard operating procedures governing working hours or budget directives involving the expenditure of funds.
- *Expected effects of use.* The use of analysis has conceptual, symbolic, and instrumental effects. Conceptual effects include the use of analysis to think about problems and solutions (conceptual use) and to legitimize preferred formulations of problems and solutions (symbolic use). By contrast, behavioral effects involve the use of policy analysis as an instrument for directing or manipulating activities or functions (instrumental use). Conceptual, symbolic, and instrumental uses of knowledge occur among individual and collective users.[87]

In practice, these three dimensions of knowledge use overlap. As we shall see in Chapter 9, the intersection among them provides a basis for assessing and improving the uses of policy analysis.

---

[85]Chester I. Barnard, *The Functions of the Executive* (Cambridge, MA: Harvard University Press, 1938, 1962), p. 198. Quoted in Carol Weiss, "Knowledge Creep and Decision Accretion," *Science Communication* 1, 3 (March 1980): 403.

[86]The concept of "ideas in good currency" is discussed by Donald A. Schon, "Generative Metaphor: A Perspective on Problem Setting in Social Policy," in *Metaphors and Thought*, ed. A. Ortony (Cambridge: Cambridge University Press, 1979), pp. 254–283.

[87]On distinctions between conceptual, instrumental, and symbolic use, see Carol H. Weiss, "Research for Policy's Sake: The Enlightenment Function of Social Science Research," *Policy Analysis* 3 (1977): 200–224; Weiss, "The Circuitry of Enlightenment," *Science Communication* 8, 2 (1986): 274–281; Nathan Caplan, Andrea Morrison, and Roger Stambaugh, *The Use of Social Science Knowledge in Policy Decisions at the National Level* (Ann Arbor, MI: Institute for Social Research, Center for the Utilization of Scientific Knowledge, 1975); Robert F. Rich, "Uses of Social Science Knowledge by Federal Bureaucrats: Knowledge for Action versus Knowledge for Understanding," in *Using Social Research in Public Policymaking,* ed. Weiss, pp. 199–211; and Karin D. Knorr, "Policymakers' Use of Social Science Knowledge: Symbolic or Instrumental?" in *Using Social Research in Public Policymaking*, ed. Carol H. Weiss, pp. 165–182.

CHAPTER SUMMARY

This chapter has presented an overview of the role of policy analysis in policymaking. Historically, the aim of policy analysis has been to provide policymakers with knowledge that can be used to solve practical problems. Policy analysis is an intellectual activity embedded in a social process. Although policymaking can be seen as a set of phases ordered in time, the organization of these phases resembles a garbage can or organized anarchy. Numerous models are available to describe how and why policies change, all of which capture an important property of policymaking.

Methods of analysis are designed to produce policy-relevant knowledge that is useful in all phases of policymaking. However, the uses of analysis are indirect, delayed, general, and ethically controversial. Although this complicates the process of knowledge use, this should be expected, considering the many variations of knowledge use based on the intersection of its composition, scope, and expected effects.

REVIEW QUESTIONS

1. What does it mean to say that policy analysis is a process of producing knowledge *of* and *in* the policymaking process?
2. What is the relation between the problems faced by societies and the growth of what has been called an "analycentric" perspective?
3. Compare and contrast intellectual and social aspects of policymaking. Give examples.
4. What is an "organized anarchy"? How is it related to the "garbage can" model of policymaking?
5. In your experience, which model(s) of policy change are most useful? Why?
6. Considering the three dimensions of knowledge use, how would you know when policymakers have used policy analysis? What would you do about it?
7. What are the pros and cons of evidence-based policymaking? Consider concepts of "technocratic guidance" and "technocratic counsel" presented in the case study for this chapter.

DEMONSTRATION EXERCISES

1. After reading Case 2.1 (Are Policy Analysts Technocrats?), evaluate the following statements about "technocratic counsel" and "technocratic guidance" in policymaking. Do you agree? Disagree? Why?
   - "The technocratic *guidance* perspective exaggerates the power and influence of professional analysts."
   - "The technocratic *counsel* perspective overestimates the symbolic importance of analysts in legitimizing policies made on political grounds."
   - "The remedy for 'technocratic counsel' " is not to abandon policy analysis, but to create a new approach called 'democratic expertise.' This approach will counteract technocratic biases "by moving from phony neutrality to thoughtful partisanship, working disproportionately to assist have-nots in understanding and making their case . . . assisting all partisans in coping with uncertainties."[88]
2. Obtain an analysis of a contemporary policy issue. The analysis could be unpublished,

[88]Rowland Egger, "The Period of Crises: 1933 to 1945," in *American Public Administration: Past, Present, and Future*, ed. Frederick C. Mosher (Tuscaloosa, AL: University of Alabama Press, 1975), pp. 49–96.

posted on a website, or published in a professional journal, magazine, or newspaper. Use Case 2.2, "Monitoring the Uses of Policy Analysis," to write a short report that answers the following questions:

- To what extent is the analysis of high *technical quality?*
- To what extent does the analysis *conform to what you expected?*
- To what extent is the analysis *actionable,* that is, has an action orientation?
- To what extent is the analysis a *challenge to the status quo?*
- What does your report suggest about the reasons why policy analysis is believed to be *useful?* Useful for what? Be specific.
- What does your report suggest about the reasons why research is actually *used*, or *not used?*

BIBLIOGRAPHY

Barber, Bernard. *Effective Social Science: Eight Cases in Economics, Political Science, and Sociology*. New York: Russell Sage Foundation, 1987.

Coalition for Evidence Based Policy, www.evidence-basedprograms.org/.

De Leon, Peter. *Advice and Consent: The Development of the Policy Sciences*. New York: Russell Sage Foundation, 1988.

Donaldson, Stewart I., Christina A. Christie, and Melvin M. Mark, eds. *Credible and Actionable Evidence: The Foundation for Rigorous and influential Evaluations*. 2nd ed. Los Angeles: Sage Publications, 2015.

Fischer, Frank. *Technocracy and the Politics of Expertise*. Newbury Park, CA: Sage Publications, 1990.

Freidson, Elliot. *Professional Powers: A Study of the Institutionalization of Formal Knowledge*. Chicago, IL: University of Chicago Press, 1986.

Horowitz, Irving L., ed. *The Use and Abuse of Social Science: Behavioral Science and Policymaking*. 2nd ed. New Brunswick, NJ: Transaction Books, 1985.

Jasanoff, Sheila. *The Fifth Branch: Science Advisors as Policymakers*. Cambridge, MA: Harvard University Press, 1990.

Kingdon, John W. *Agendas, Alternatives, and Public Policies*. Glenview, IL: Scott, Foresman, 1984.

Lerner, Daniel, ed. The *Human Meaning of the Social Sciences*. Cleveland, OH: World Publishing Company, 1959.

Lindblom, Charles E. *Inquiry and Change: The Troubled Attempt to Understand and Change Society*. New Haven, CT: Yale University Press, 1990.

Lindblom, Charles E., and David K. Cohen. *Usable Knowledge: Social Science and Social Problem Solving*. New Haven, CT: Yale University Press, 1979.

Lindblom, Charles E., and Edward J. Woodhouse. *The Policy-Making Process*. 3rd ed. Englewood Cliffs, NJ: Prentice-Hall, 1993.

Machlup, Fritz. *Knowledge: Its Creation, Distribution, and Economic Significance*. Vol. 1 of Knowledge and Knowledge Production. Princeton, NJ: Princeton University Press, 1980.

MacRae, Duncan, Jr. *The Social Function of Social Science*. New Haven, CT: Yale University Press, 1976.

Parsons, Wayne. "From Muddling Through to Muddling Up: Evidence Based Policy-Making and the Modernisation of British Government." Unpublished paper. London: University of London, 2004.

Parsons, Wayne. *Public Policy: An Introduction to the Theory and Practice of Policy Analysis*. Northampton, MA: Edward Elgar Publishing Inc., 1995.

Ravetz, Jerome. *Science and Its Social Problems*. Oxford: Oxford University Press, 1971.

Sabatier, Paul A. *Theories of the Policy Process*. Boulder, CO: Westview Press, 1999.

Schmandt, Jurgen, and James E. Katz. "The Scientific State: A Theory with Hypotheses." *Science, Technology, and Human Values* 11 (1986): 40–50.

United Kingdom, House of Commons, Science and Technology Committee. *Scientific Advice, Risk and Evidence Based Policy Making. Seventh Report of Session 2005–06*, Volume I. London: HMO Printing House, 2006.

Weiss, Carol H. *Social Science Research and Decision Making*. New York: Columbia University Press, 1980.

CASE 2.1 ARE POLICY ANALYSTS TECHNOCRATS?

In the first half of the twentieth century, the involvement of social and natural scientists in policymaking was largely a response to social, economic, and politico-military crises.[89] Members of the sciences and professions were ad hoc and irregular participants in government in the period of depression, dislocation, and war that stretched from the early 1930s to the late 1940s. In succeeding years, however, the social sciences and social professions—political science, sociology, economics, public administration, business management, social work, planning—became one of the main sources of full-time governmental staff, advisers, and consultants. For the first time, public agencies recruited on a regular basis specialists who had received their training and had been certified by their respective professional organizations.

For many observers, this new era signified the advent of a new form of social and political organization in which public policymaking and societal guidance were critically dependent on specialized knowledge and technologies of the sciences and professions. Emphasizing its social aspect, some referred to this new form of organization as the "knowledge society"[90] or "postindustrial society."[91] Others, focusing on its political aspect, spoke of the transition from the "administrative state" to the "scientific state."[92]

Postindustrial Society and the Scientific State

Postindustrial society, an extension of patterns of policymaking and socioeconomic organization of industrial society, is increasingly dominated by an educated professional-technical class. Many of the characteristics of a postindustrial society are important to consider when we ponder the historical evolution and significance of policy analysis:[93]

- *Centrality of theoretical knowledge.* Although the earliest civilizations relied on specialized knowledge, only in the late twentieth and twenty-first centuries are innovations in "soft" (social) and "hard" (physical) technologies directly dependent on the codification of theoretical knowledge provided by the social, physical, and biological sciences.
- *Creation of new intellectual technologies.* The improvement in mathematical and statistical techniques has made it possible to use advanced modeling, simulation, and various forms of systems analysis to find more efficient and "rational" solutions to public problems.
- *Spread of a knowledge class.* The most rapidly expanding occupational group in the United States is composed of technicians and professionals. This group, including professional managers, represented 28 percent of employed persons in 1975, 28.3 percent in 1995, and 30.3 percent in 2000. Persons in nonfarm occupations grew from 62.5 to 98.4 percent between 1900 and 1990.
- *Shift from goods to services.* In 1975, more than 65 percent of the active labor force was engaged in the production and delivery of services, a figure that exceeded 80 percent by 2000. Services are primarily human (e.g., health, education, social welfare) and professional-technical (e.g., knowledge services, computer programming, telecommunications).

(*continued*)

[89]Rowland Egger, "The Period of Crises: 1933 to 1945," in *American Public Administration: Past, Present, and Future*, ed. Frederick C. Mosher (Tuscaloosa, AL: University of Alabama Press, 1975), pp. 49–96.
[90]Fritz Machlup, *The Production and Distribution of Knowledge in the United States* (Princeton, NJ: Princeton University Press, 1962).
[91]Daniel Bell, *The Coming of Post-Industrial Society: A Venture in Social Forecasting* (New York: Basic Books, 1976).
[92]Jurgen Schmandt and James E. Katz, "The Scientific State: A Theory with Hypotheses," *Science, Technology, and Human Values* 11 (1986): 40–50.
[93]Bell, *Coming of Post-Industrial Society*, pp. xvi–xviii.

- *Instrumentalization of sciences.* Although the natural and social sciences have long been used as instruments for controlling the human and material environment, they have been largely open to internal criticism and protected by norms of scientific inquiry. In the present period, science and technology have become increasingly bureaucratized, subordinated to government goals, and assessed in terms of their payoffs to society. More than 95 percent of all funded research is applied research on practical problems of business, commerce, and government.
- *Scarcity of scientific and technical knowledge.* Knowledge has increasingly become one of society's most scarce resources. Knowledge is a collective, not a private good, and cooperative rather than competitive strategies for its production and use are required if optimal results are to be achieved. The rapid exponential growth of the Internet and World Wide Web are vehicles for satisfying the demand for knowledge by governments, businesses, and nonprofit organizations.

The development of policy analysis and other applied sciences seems to be part of an emerging postindustrial society and scientific state. Yet, what does this development signify for issues of democratic governance, civil liberties, and privacy? How much influence do producers of policy-relevant knowledge have? Although theoretical knowledge clearly seems to be more central than it has been in other historical periods, what does this mean for the distribution of political power? Does the proliferation of new intellectual technologies mean also that structures and processes of policymaking have changed? Is the spread of a "knowledge class" commensurate with its power and influence? Considering that the aim of policy analysis and other applied sciences is to produce knowledge for practical purposes, whose purposes are being served? In short, how are we to interpret the historical transformation of policy-oriented inquiry from earlier times to the present-day "knowledge society" and "scientific state"?

Technocratic Guidance versus Technocratic Counsel

There are several contending perspectives of these questions, two of which we consider here.[94] One perspective holds that the professionalization of policy analysis and other applied social sciences is—or should be—shifting power and responsibility from policymakers to policy analysts.[95] This perspective, *technocratic guidance,* is associated with the "analycentric turn" described earlier in this chapter. The basic premise of the analycentric turn is that "the surest way to improve the quality of public choice is to have more analysts producing more analyses."[96] By contrast, a contending perspective, *technocratic counsel,* holds that the professionalization of policy analysis and other applied social sciences signifies new and more effective ways to enhance the influence of policymakers and other dominant groups whose positions continue to rest on power, wealth, and privilege.[97]

Strictly speaking, neither of these perspectives is wholly accurate in its interpretation of events surrounding the movement toward postindustrial society; each contains a one-sided emphasis on particular characteristics of contemporary society to the exclusion of others. Yet just as the development of nineteenth-century policy analysis was a practical response to problems of the day as viewed by dominant groups, so too is contemporary policy analysis a consequence of changes in the structure and role of government as it has attempted to grapple with new problems. "The contemporary boom in policy

[94]See Jeffrey D. Straussman, "Technocratic Counsel and Societal Guidance," in *Politics and the Future of Industrial Society*, ed. Leon N. Lindberg (New York: David McKay, 1976), pp. 126–166.
[95]For example, Amitai Etzioni, *The Active Society* (New York: Free Press, 1968), and Donald T. Campbell, "The Experimenting Society," in *The Experimenting Society: Essays in Honor of Donald T. Campbell*, ed. William N. Dunn (New Brunswick, NJ: Transaction Books, 1998), pp. 35–68.
[96]Allen Schick, "Beyond Analysis," *Public Administration Review* 37, 3 (1977): 259.
[97]See, for example, Guy Benveniste, *The Politics of Expertise* (Berkeley, CA: Glendessary Press, 1972), and Frank Fischer, *Technocracy and the Politics of Expertise* (Newbury Park, CA: Sage Publications, 1990).

analysis," writes Schick, "has its primary source in the huge growth of American governments, not in the intellectual development of the social sciences."[98] Throughout the twentieth century, the expansion of government was followed by demands for more policy-relevant knowledge. As government has grown, so too has the market for policy analysis.[99]

The development of policy analysis in this century has followed fits and starts in the growth of federal executive agencies.[100] In 1900, there were some ninety federal executive agencies. By 1940, this figure had grown to 196, and, by 1973, the total number of federal executive agencies had jumped to 394. The greatest percentage increase occurred after 1960. Some 35 percent were created after 1960. The rapid growth of the federal government was mainly a response to new problems in areas of national defense, transportation, housing, health, education, welfare, crime, energy, and the environment. There were many new "challenges" to policymaking:[101] participation overload, mobilization of service sector workers, transformation of basic social values, realignment of interest groups, growing cleavages between urban and suburban citizens, and recurrent fiscal crises. Some observers, notably policy scientist Yehezkel Dror, went so far as to formulate the relationship between such problems and policy analysis in the form of a general law: "While the difficulties and dangers of problems tend to increase at a geometric rate, the number of persons qualified to deal with these problems tends to increase at an arithmetic rate."[102] Thus, the growth of policy analysis seems to be a consequence, not a cause, of changes in the structure of government and the nature and scope of social problems.

The *technocratic guidance* perspective asserts that policy-relevant knowledge is an increasingly scarce resource whose possession enhances the power and influence of policy analysts. The technocratic guidance perspective may be summarized in terms of five major propositions:[103]

- The growing interdependencies, complexity, and pace of change of contemporary society make existing knowledge obsolete, thus increasing the demand for new forms of policy-relevant knowledge.
- Problems of contemporary society may be resolved with specialized knowledge produced by policy analysts.
- The technical complexity of policy choices prompts higher levels of direct involvement of policy analysts and other applied scientists.
- Higher levels of direct involvement enhance the power of professional analysts to influence policymaking, making politicians increasingly dependent on them.
- The growing dependence of politicians on professional policy analysts erodes their political power.

A rival perspective, that of *technocratic counsel,* begins from the assumption that professional policy analysts operate in settings in which policymakers, as the consumers of specialized knowledge, determine to a large extent the activities of producers. In this context, the primary role of analysts is to legitimize—that is, justify in scientific and technical terms—policy decisions made by the real holders of power. The technocratic counsel perspective may also be summarized in terms of several key propositions.[104]

- Major policy alternatives reflect conflicting values held by different segments of the community.
- Value conflicts are associated with disparities of political power.
- The choice of a given policy alternative symbolizes the victory of one segment of the community over another.

(*continued*)

[98]Schick, "Beyond Analysis," p. 258.
[99]Schmandt and Katz, "The Scientific State."
[100]Herbert Kaufman, *Are Government Organizations Immortal?* (Washington, DC: Brookings Institution, 1976), pp. 34–63.
[101]Samuel P. Huntington, "Postindustrial Politics; How Benign Will It Be?" *Comparative Politics* 6 (January 1974): 163–192; Ronald Inglehart, "The Silent Revolution in Europe: Intergenerational Change in Postindustrial Societies," *American Political Science Review,* 65 (December 1971): 991–1017; and James O'Connor, *The Fiscal Crisis of the State* (New York: St. Martin's Press, 1973).
[102]Yehezkel Dror, *Ventures in Policy Sciences: Concepts and Applications* (New York: American Elsevier, 1971), p. 2.
[103]Straussman, "Technocratic Counsel," pp. 150–151.
[104]Ibid., pp. 151–152.

- Policymakers use scientific and technical justifications produced by analysts to suppress conflicts and legitimize choices that are actually made on political grounds.
- The credibility of scientific and technical justifications requires that policy analysis and other applied sciences be presented as value neutral, impartial, and apolitical.
- Professional analysts, although they are a source of scientific and technical justifications, are expendable. They also serve as convenient scapegoats for policies that fail.

If we pick our cases carefully, we can show that analysts exercise a large measure of "technocratic guidance." For example, the New York City–RAND Corporation's analysis of factors influencing the response time of city firefighters to reported fires has been credited with unusual success, although other RAND efforts have been less effective in shaping decisions.[105] On the other hand, the use of specialized knowledge to make policy choices appears to be highly uneven. For example, in a study of 204 federal executives conducted in 1973–1974, social science research was relatively unimportant as a basis for choosing among policy alternatives.[106] Moreover, the extent to which research was used depended on essentially nontechnical factors: "the level of knowledge utilization is not so much the result of a slow flow of relevant and valid knowledge from knowledge producers to policymakers, but is due more to factors involving values, ideology, and decision-making styles."[107]

The finding that policy analysis is used for political purposes adds a measure of credibility to the technocratic counsel perspective. So does the apparent conservative character of many analyses. In this context, policy analysis has been characterized as a conservative and superficial kind of social science that fails to pose radical questions about basic social values and institutions, and neglects policy alternatives that depart significantly from existing practices.[108] Moreover, some observers see in policy analysis an ideology in disguise that suppresses ethical and value questions in the name of science.[109] Under such conditions, it is understandable that analysts might be used as instruments of everyday politics. ■

[105]On this and other cases, see Greenberger, Crenson, and Crissey, *Models in the Policy Process,* pp. 231–318.

[106]Nathan Caplan et al., *The Use of Social Science Knowledge in Policy Decisions at the National Level* (Ann Arbor, MI: Institute for Social Research, Center for Research on the Utilization of Scientific Knowledge, 1975).

[107]Nathan Caplan, "Factors Associated with Knowledge Use among Federal Executives," *Policy Studies Journal* 4, 3 (1976): 233. See also Robert F. Rich, *Social Science Knowledge and Public Policymaking* (San Francisco, CA: Jossey-Bass, 1981), and David J. Webber, "The Production and Use of Knowledge in the Policy Process," in *Advances in Policy Studies since 1950*, ed. Dunn and Kelly, pp. 415–441.

[108]Charles E. Lindblom, "Integration of Economics and the Other Social Sciences through Policy Analysis," in *Integration of the Social Sciences through Policy Analysis,* ed. James C. Charlesworth (Philadelphia: American Academy of Political and Social Sciences, 1972), p. 1. For an elaboration of this critique, see Charles E. Lindblom, *Inquiry and Change: The Troubled Attempt to Understand and Change Society* (New Haven, CT: Yale University Press, 1990).

[109]Laurence H. Tribe, "Policy Science: Analysis or Ideology?" *Philosophy and Public Affairs* 2, 1 (1972): 66–110.

CASE 2.2 MONITORING THE USES OF POLICY ANALYSIS

The following profile, based on the work of Carol H. Weiss, is helpful in monitoring the uses of policy analysis.[110] Obtain a report from a website such as www.publicagenda.org, www.rand.org, www.aei.org, or a similar source. Please use the following scale for responses to questions about the report.[111]

1 = very much
2 = quite a bit
3 = to some extent
4 = not much
5 = not at all

1. To what extent does the analysis contain knowledge that is relevant to your work?
2. To what extent do you think the analysis is reliable (could be repeated with similar conclusions) and valid (plausibly explains what it set out to explain)?
3. To what extent do the conclusions agree with your own views about the issue?
4. Please indicate the extent to which the analysis
   a. deals with a high priority issue (AO)
   b. adds to theoretical knowledge in the field (TQ)
   c. adds to practical knowledge about policies or programs (AO)
   d. agrees with your ideas or values (CUE)
   e. identifies outcomes that policymakers can do something about (AO)
   f. suggests potential courses of action (AO)
   g. implies the need for major policy changes (CSQ)
   h. focuses on specific policy outcomes (AO)
   i. contains explicit recommendations for action (AO)
   j. supports your own position on the issue (CUE)
   k. provides evidence to back up recommendations (TQ)
   l. contains politically feasible implications (AO)
   m. agrees with previous knowledge about the issue (TQ)
   n. can be implemented under the present circumstances (AO)
   o. challenges existing assumptions and institutional arrangements (CSQ)
   p. raises new issues or perspectives (CSQ)
   q. is inexpensive to implement (AO)
   r. was available at the time a decision had to be made (AO)
   s. contains novel or unexpected findings (negative coding)
   t. provides quantitative data (TQ)
   u. can be generalized to similar settings or persons (TQ)
   v. addresses in a thorough manner different factors that could explain outcomes (TQ)
   w. uses statistically sophisticated analyses (TQ)
   x. demonstrates high technical competence (TQ)
   y. presents internally consistent and unambiguous findings (TQ)
   z. takes an objective and unbiased approach (TQ)

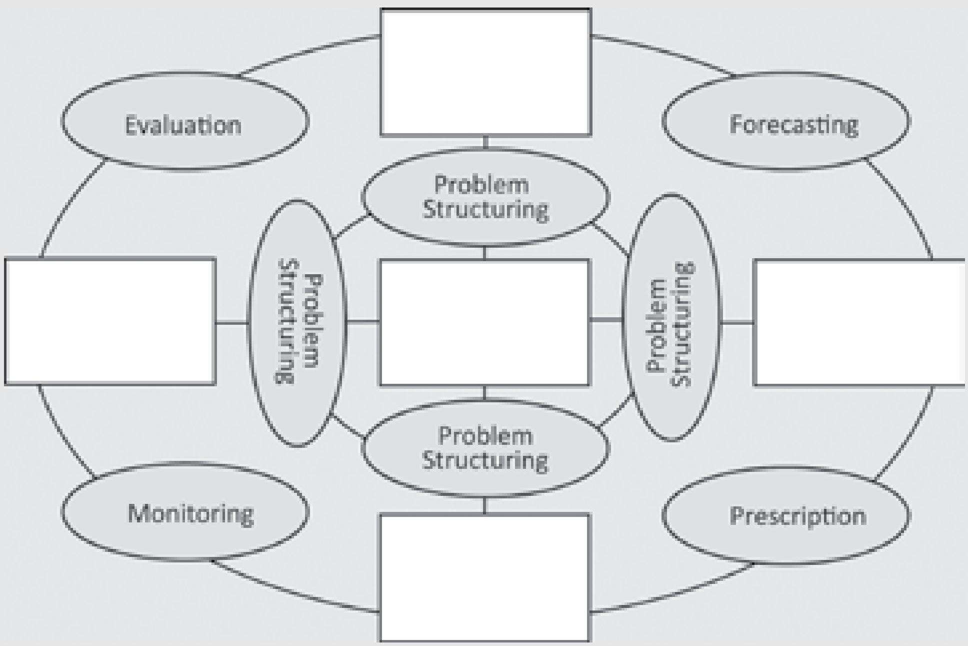
Use the following categories to sum the scaled responses (numbers 1 through 5), and then average them (divide by the number of questions answered). The capital letters after each question (e.g., TQ) refer to the categories.

Technical Quality (TQ):
Conformity to User Expectations (CUE):
Action Orientation (AO):
Challenge Status Quo (CSQ): ■

[110] These questions are drawn from Carol H. Weiss and Michael J. Bucuvalas, "The Challenge of Social Research to Decision Making," in *Using Social Research in Public Policymaking,* (Lexington, MA: D. C. Heath, 1977), Appendix 15A and 15B, pp. 231–233.
[111] I am grateful to Carol Weiss for permitting me to use these questions and, in some cases, suggesting that they be modified.

PART II

Methods of Policy Analysis



CHAPTER 3

Structuring Policy Problems

LEARNING OBJECTIVES

By studying this chapter, you should be able to:

- Distinguish problem structuring from problem solving
- Understand the subjective, systemic, and interdependent nature of problems
- Contrast problem situations and problems
- Distinguish relatively well-structured, moderately structured, and ill-structured problems
- Contrast tame problems and wicked problems
- Compare and contrast types of policy models
- Assess strengths and limitations of different methods of problem structuring
- Use a Pareto chart to estimate the boundary of a relatively ill-structured problem

INTRODUCTION

Inexperienced analysts suppose that facts speak for themselves. This epistemologically innocent view fails to recognize that the same facts are interpreted in varied ways. Identical policy-relevant information may evoke conflicting definitions of a problem, leading to different interpretations of the same facts. These interpretations are shaped by assumptions about human nature, the proper role of government, the nature of the economy, and the definition of knowledge itself.

NATURE OF POLICY PROBLEMS

Policy problems are unrealized needs, values, or opportunities for improvement.[1] Information about the nature, scope, and severity of a problem is produced by applying the policy-analytic procedure of problem structuring. Problem structuring, which is a phase of inquiry in which analysts compare, contrast, and evaluate competing formulations of a problem, is among the most important procedures performed by analysts. Problem structuring is a central guidance system that affects the success of other phases of policy analysis. Analysts seem to fail more often because they solve the wrong problem than because they get the wrong solution to the right problem.[2]

Problem Solving versus Problem Structuring

Policy analysis is often described as a problem-solving methodology. Although this is partly correct—and analysts do succeed in finding solutions for problems[3]—the problem-solving image is incomplete. The problem-solving image wrongly suggests that analysts can successfully identify, evaluate, and recommend solutions for a problem without investing considerable time in formulating that problem. Policy analysis is best seen as a dynamic, multilevel process in which methods of problem structuring are essential to the success of methods of problem solving (Figure 3.1).

Methods of problem structuring are different from methods of problem solving because the questions they are designed to answer are different. Methods of problem solving are more appropriate and effective when we know more or less precisely what problem to solve. Problems involving the calculation of the net benefits of programs for reducing industrial pollution are on the whole precisely formulated. Problems involving judgments about the scope and severity of industrial pollution and the identification of conditions that may cause it are imprecisely formulated and poorly defined. Here we may find that the proper formulation of the problem is closely related to the driving habits of Americans for whom petroleum fuels are comparatively cheap, by world standards, and subsidized by the government. This is a question requiring problem-structuring methods.

[1]See David Dery, *Problem Definition in Policy Analysis* (Lawrence: University Press of Kansas, 1984).
[2]Russell A. Ackoff, *Redesigning the Future: A Systems Approach to Societal Problems* (New York: Wiley, 1974), p. 21.
[3]See, for example, Bernard Barber, *Effective Social Science: Eight Cases in Economics, Political Science, and Sociology* (New York: Russell Sage Foundation, 1987).

The more or less precisely formulated problem involving calculation is a question of problem solving. Just as we may distinguish problem solving and problem structuring, we may also contrast:

- *Problem sensing versus problem structuring.* The process of problem solving rarely begins with clearly articulated problems, but with a sense of diffuse worries and inchoate signs of stress.[4] These diffuse worries and inchoate signs of stress are not problems but *problem situations,* which refer to the state of "unsettled belief" we face when we sense vague and ill-defined situations requiring action. By contrast, *policy problems* are products of applying problem-structuring methods to problem situations. Problems are elements of problem situations that have been abstracted from these situations through problem structuring."[5]
- *Problem structuring versus problem solving.* Policy analysis includes methods of *problem structuring* as well as methods of *problem solving.* Problem-structuring methods provide a methodological complement to theories of policy design, or design science.[6] Problem-structuring methods are *metamethods*—that is, they are "about" and "come before" methods of problem solving.
- *Problem resolving* versus *problem unsolving and problem dissolving.* The terms *problem resolving*, *problem unsolving*, and *problem dissolving* refer to three types of error correction.[7] Although the three terms come from the same root (L. *solvere*, to solve or dissolve), the error-correcting processes to which they refer are different (Figure 3.1). *Problem resolving* involves the reanalysis of a correctly structured problem to reduce calibrational errors, for example, reducing the probability of Type I or Type II errors in testing the null hypothesis that a policy has no effect. *Problem unsolving,* by contrast, involves the abandonment of a solution based on the wrong formulation of a problem—for example, the policy of urban renewal implemented in central cities during the 1960s—and a return to problem structuring in an attempt to formulate the right problem. Finally, *problem dissolving* involves the abandonment of an incorrectly formulated problem and a return to problem structuring to start the process anew.

---

[4]Martin Rein and Sheldon H. White, "Policy Research: Belief and Doubt," *Policy Analysis* 3, 2 (1977): 262. This is a paraphrase of John Dewey.

[5]Russell A. Ackoff, *Redesigning the Future: A Systems Approach to Societal Problems* (New York: Wiley, 1974), p. 21. Ackoff does not specify that the process of abstraction occurs through problem structuring, using the more general term "analysis."

[6]See Stephen H. Linder and B. Guy Peters, "From Social Theory to Policy Design," *Journal of Public Policy* 4, no. 4 (1985): 237–259; John Dryzek, "Don't Toss Coins into Garbage Cans: A Prologue to Policy Design," *Journal of Public Policy* 3, 3 (1983): 345–367; and Trudi C. Miller, "Conclusion: A Design Science Perspective," in *Public Sector Performance:* A *Turning Point,* ed. T. C. Miller (Baltimore: Johns Hopkins University Press, 1985).

[7]Russell L. Ackoff, "Beyond Problem Solving," *General Systems* 19 (1974): 237–239.

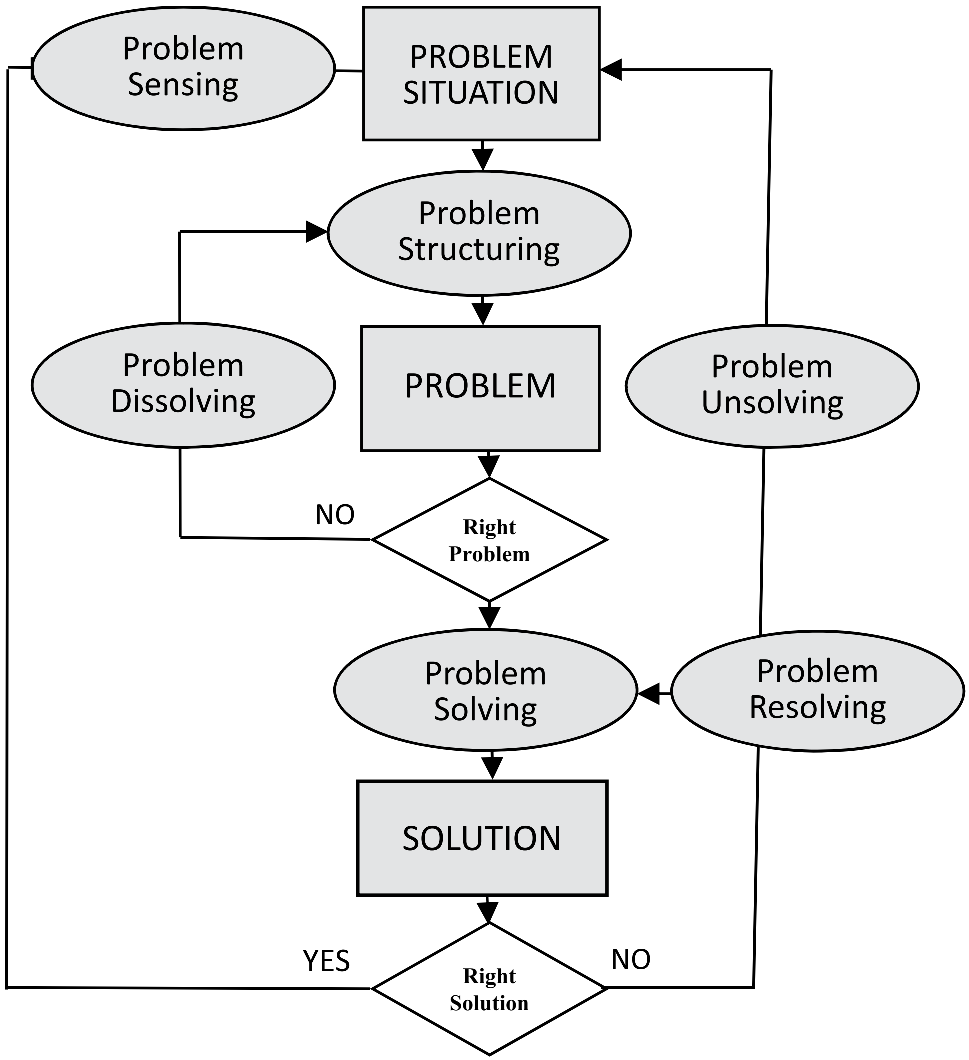


FIGURE 3.1
The Process of Problem Structuring

Source: Dunn (1988): 720–737.

CHARACTERISTICS OF PROBLEMS

There are several important characteristics of policy problems:

1. *Interdependency.* Policy problems in one area (e.g., energy) frequently affect policy problems in other areas (e.g., health care and unemployment). In reality, policy problems are not independent entities; they are parts of whole systems of problems

best described as *messes,* that is, systems of external conditions that produce dissatisfaction among different segments of the community.[8] Systems of problems (messes) are difficult or impossible to resolve by using a *reductionistic approach*—that is, one that decomposes problems into their component elements or parts—because rarely can problems be defined and resolved independently of one another. Sometimes it is easier "to solve ten interlocking problems simultaneously than to solve one by itself."[9] Systems of interdependent problems require a *holistic approach,* that is, one that views problems as inseparable and unmeasurable apart from the system of which they are interlocking parts.[10]

2. *Subjectivity.* The external conditions that give rise to a problem are selectively defined, classified, and evaluated. Although there is a sense in which problems are objective—for example, air pollution may be defined in terms of levels of CO_2 gases and particulates in the atmosphere—the same data about pollution is interpreted in markedly different ways. The reason is that policy problems are abstracted from problem situations by analysts. Problems are partly subjective and partly objective, a characterization supported by the epistemological doctrine of *objective relativism*. Problems are relative but nevertheless objective.[11]
3. *Artificiality.* Policy problems are possible only when human beings make judgments about the desirability of altering some problem situation. Policy problems are products of subjective human judgment; policy problems come to be accepted as legitimate definitions of objective social conditions; policy problems are socially constructed, maintained, and changed.[12] Problems have no existence apart from the individuals and groups who define them, which means that there are no "natural" states of society that in and of themselves constitute policy problems.
4. *Instability.* There may be as many different solutions for a problem as there are definitions of that problem. "Problems and solutions are in constant flux; hence problems *do not stay solved*. . . . Solutions to problems become obsolete even if the problems to which they are addressed do not."[13]

Problems are not discrete mechanical entities; they are purposeful (teleological) systems in which no two members are identical in all or even any of their properties or behaviors; the properties or behavior of each member has an effect on the properties or behavior of the system as a whole; the properties and behavior of each member, and the way each affects the system as a whole, depend on the properties and behavior of at least one other member of the system; and all possible subgroups of members have a non-independent effect on the system

[8]Russell L. Ackoff, *Redesigning the Future: A Systems Approach to Societal Problems* (New York: Wiley, 1974), p. 21.
[9]Harrison Brown, "Scenario for an American Renaissance," *Saturday Review,* December 25, 1971, 18–19.
[10]See Ian I. Mitroff and L. Vaughan Blankenship, "On the Methodology of the Holistic Experiment: An Approach to the Conceptualization of Large-Scale Social Experiments," *Technological Forecasting and Social Change* 4 (1973): 339–353.
[11]Ackoff, *Redesigning the Future,* p. 21; Abraham Kaplan, *The Conduct of Inquiry: Methodology for Behavioral Science* (San Francisco, CA: Chandler, 1964), pp. 392–393
[12]Peter L. Berger and Thomas Luckmann, *The Social Construction of Reality,* 2nd ed. (New York: Irvington, 1980).
[13]Ackoff, *Redesigning the Future*, p. 21.

as a whole.[14] Subjectively experienced problems—crime, poverty, unemployment, inflation, energy, pollution, health, security—cannot be decomposed into independent subsets without running the risk of producing an approximately right solution to the wrong problem.

A key characteristic of systems of problems is that the whole is greater—that is, qualitatively different—than the simple sum of its parts. A pile of stones may be defined as the sum of all individual stones but also as a pyramid. Similarly, a human being can think and act, but none of its parts can do so independently. Membership in a system may increase or diminish the capabilities of each element, but it does not leave them unaffected. Clearly, a brain that is not part of a living body cannot function. "An individual who is part of a nation or a corporation is thereby precluded from doing some things he could otherwise do, and he is enabled to do others he could not otherwise do."[15]

Finally, recognition of the interdependency, subjectivity, artificiality, and instability of problems alerts us to the possible *unanticipated consequences* that follow from policies that are based on the right solution to the wrong problem. Consider, for example, the problem situation confronted by Western European governments in the late 1970s. France and West Germany, seeking to expand the supply of available energy by constructing nuclear power plants on the Rhine River, defined the energy problem in a way that assumed that the production of nuclear power is independent of other problems. Ivan Illich went so far as to predict that malaria would become a major epidemic in Europe within 10 years, because Germany and France had decided "to build atomic generators that utilize river waters for their cooling systems and hence bring the water temperature within the range in which anopheles (the malaria-carrying mosquito) breeds.[16]

Although the forecast proved to be incorrect, the understanding that energy policy involves a system of problems was not.

Problems versus Issues

Policy issues involve disagreements about actual or potential courses of action. Issues also reflect competing views of the nature of problems themselves. An apparently clear-cut issue—for example, whether the government should impose air quality standards on industry—is typically the consequence of conflicting sets of assumptions about the nature of pollution:[17]

1. Pollution is a natural consequence of capitalism, an economic system where the owners of industry naturally seek to increase profits from their investments. Some damage to the environment is a necessary price to pay for a healthy capitalist economy.
2. Pollution is a result of the need for power and prestige among industrial managers who seek promotions in large career-oriented bureaucracies. Pollution has been just as severe in socialist systems where there are no profit-seeking private owners.

---

[14]Mitroff and Blankenship, "Methodology of the Holistic Experiment," pp. 341–342.

[15]Ackoff, *Redesigning the Future*, p. 13.

[16]Ivan Illich in conversation with Sam Keen, reported in *Psychology Today* (May 1976).

[17]See Ritchie P. Lowry, *Social Problems: A Critical Analysis of Theories and Public Policy* (Lexington, MA: D. C. Heath and Company, 1974), pp. 23–25.

3. Pollution is a consequence of consumer preferences in high mass-consumption society. In order to ensure corporate survival, owners and managers must satisfy consumer preferences for high-performance engines and automobile travel.

The ability to recognize differences between problem situations and policy problems is important for understanding the ways that common experiences are translated into policy issues. The formulation of a problem is influenced by the assumptions that policy stakeholders bring to a problem situation. In turn, different formulations of the problem shape the ways that policy issues are defined. In our earlier example of environmental pollution, assumptions about the operation of a healthy capitalist economy may result in a negative view of government regulation of air quality standards in industry, whereas assumptions about corporate managerial behavior may result in an affirmative position. By contrast, assumptions about consumer preferences and corporate survival may influence the view that government regulation is a nonissue, because government cannot legislate consumer demand.

Policy issues may be classified according to a hierarchy of types: major, secondary, functional, and minor (Figure 3.2). *Major issues* are those encountered at the highest levels of government within and between national, regional, and local jurisdictions. Major issues typically involve questions of agency mission, that is, questions about the nature and purpose of an organization. The issue of whether the purpose of the Department of Health and Human Services is the elimination of conditions that give rise to poverty is a question of agency mission. *Secondary issues* are those located at the level of agency programs at the federal, state, and local levels. Secondary issues involve setting program priorities and the definition of target groups and beneficiaries.

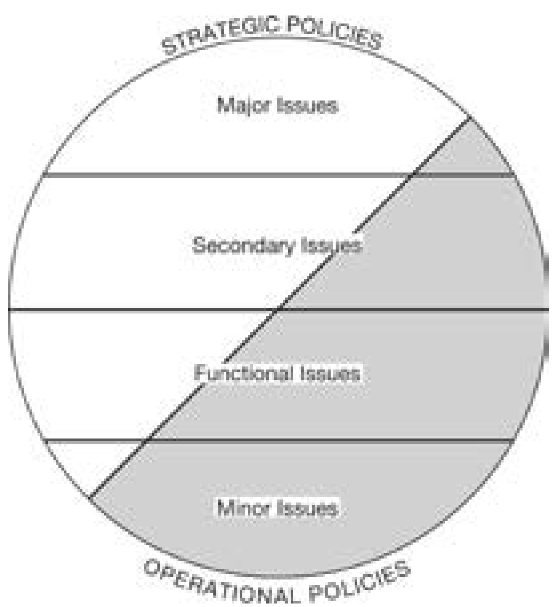


FIGURE 3.2
Hierarchy of Types of Policy Issues

The issue of how to define poverty families is a secondary issue. *Functional issues,* by contrast, are those located at the program and the project levels involving questions of budgeting, finance, and procurement. Finally, *minor issues* involve staffing, employee benefits, vacation times, working hours, and standard operating procedures and rules.

Some types of issues are strategic, whereas others are operational. A *strategic policy* is one where the consequences of decisions are relatively irreversible. Such issues as whether the United States should send troops to Iraq or Afghanistan, or whether there should be universal health care, call for strategic policies involving actions that cannot be reversed for many years. By contrast, *operational policies*—that is, policies where the consequences of decisions are relatively reversible—do not involve the risks and uncertainty present at higher levels. Although all types of policies are interdependent, and therefore cross levels, the complexity and irreversibility of policies increases as one moves up the hierarchy of types of policy issues.

Three Classes of Policy Problems

There are three classes of policy problems: well-structured, moderately structured, and ill-structured problems.[18] The classification of a problem is determined by its relative complexity. The differences among well-structured, moderately structured, and ill-structured problems may be illustrated by considering properties of their respective elements (Table 3.1).

Well-structured problems involve one or a few decision-makers and a small set of alternatives. There is consensus on utilities (values), which are clearly ranked in order of decision-makers' preferences. The outcomes of each alternative are known either with complete certainty (deterministically) or within an acceptable margin of error. The prototype of the well-structured problem is the algorithmic decision problem, where all consequences of all policy alternatives are programmed by a computer in advance. Operational problems in public agencies faced with replacing agency vehicles are relatively simple well-structured problems that involve finding the optimum point at which an older vehicle should be exchanged for a new one, taking into account average repair costs for older vehicles and purchasing and depreciation costs for new ones.

Moderately structured problems involve a few decision-makers and a relatively limited number of alternatives. Disagreements about utilities (values) are manageable and may be resolved by bargaining and mutual adjustment. Nevertheless, the outcomes of alternatives are uncertain. The prototype of the moderately structured problem is the simulation or game, for example, the "prisoner's dilemma."[19] In this game, two prisoners are held in separate cells, where each is interrogated by the prosecuting attorney, who must obtain

[18]The first extensive discussion of ill-structured problems (which he called ill-defined) was by Walter R. Reitman, "Heuristic Decision Procedures, Open Constraints, and the Structure of Ill-Defined Problems," *Human Judgments and Optimality* (1964): 282–315; and Herbert A. Simon, "The Structure of Ill Structured Problems." *Artificial Intelligence* 4, 3–4 (1973): 181–201. See also Francisco Sagasti, "Epistemology as General Systems Theory: An Approach to the Design of Complex Decision-Making Experiments," *Philosophy of the Social Sciences* 3 (1973): 117–134.

[19]Anatol Rapoport and Albert M. Chammah, *Prisoner's Dilemma* (Ann Arbor: University of Michigan Press, 1965).

TABLE 3.1

Differences in the Structure of Three Classes of Policy Problems

	Problem Structure		
Element	*Well Structured*	*Moderately Structured*	*Ill Structured*
Decision-maker(s)	One	Few	Many
Alternatives	Fixed	Limited	Unlimited
Utilities (values)	Consensus	Bargaining	Conflict
Outcomes	Certain	Uncertain	Risky
Probabilities	Deterministic	Estimable	Estimable

a confession from one or both prisoners to obtain a conviction. The prosecutor, who has enough evidence to convict each prisoner of a lesser crime, tells each that if neither confesses, they will both be tried for a lesser crime carrying lesser punishment; if both confess to the more serious crime, they will both receive a reduced sentence; but if only one confesses, he will receive probation, whereas the other will receive a maximum sentence. The "optimal" choice for each prisoner, given that neither can predict the outcome of the other's decision, is to confess. Yet it is precisely this choice that will result in 5-year sentences for both prisoners, because both attempt to minimize their sentences. This example not only illustrates the difficulties of making decisions when outcomes are uncertain but also shows that otherwise "rational" individual choices may contribute to collective irrationality in small groups, agencies, and corporations.

Ill-structured problems involve many decision-makers whose utilities (values) are either completely unknown or, if known, impossible to rank in a consistent fashion. Whereas moderately structured problems may be resolved through bargaining, a major characteristic of ill-structured problems is conflict among competing goals. Alternatives and outcomes may also be unknown, or subject to risks that may be estimated probabilistically. Here, the problem is not to uncover known deterministic relations, but rather to define the nature of the problem. The prototype of the ill-structured problem is the completely intransitive decision problem, that is, one where it is impossible to select a single policy alternative that is preferred to all others. Whereas well-structured and moderately structured problems contain preference rankings that are *transitive*—that is, if alternative A_1 is preferred to alternative A_2, and alternative A_2 is preferred to alternative A_3, then alternative A_1 is preferred to alternative A_3—ill-structured problems have intransitive preference rankings.

Whether faced by an individual, a small group, or an organization, many important problems are relatively ill-structured. Although some have proposed that we use the terms "wicked" and "tame" problems in place of ill-structured and well-structured problems, the wicked-tame distinction appears to have been created for primarily rhetorical purposes.[20]

[20]In 1967, C. West Churchman designated such problems as "wicked." C. West Churchman. "Guest Editorial: Wicked Problems," *Management Science* 14, 4 (1967): B141–B142. Horst Rittel and other colleagues in planning

For this reason, we retain the distinction between relatively well-structured, moderately structured, and ill-structured problems. The qualifier "relatively" reflects the proviso that the distinctions mark approximate points on a continuum.

The reasons why ill-structured problems are important for public policy have been ably summarized by a number of social scientists. Relatively ill-structured problems are important because in policy contexts it is unrealistic to assume the existence of one or a few decision-makers with consistent (or transitive) preferences and consensus on objectives, because policymaking typically involves bargaining, competition, and conflict.[21] Moreover, it is seldom possible to identify the full range of alternative solutions for a problem, in part because of constraints on the acquisition of information, but also because it is difficult to know when we have collected enough information to know when a solution has been found. Rittel and Webber state this concisely when they propose that there is no "stopping rule" to determine when a problem has been solved.[22]

PROBLEM STRUCTURING IN POLICY ANALYSIS

Methods for solving ill-structured problems are not the same as those for solving well-structured problems.[23] Whereas well-structured problems permit the use of more or less standard methods, ill-structured problems demand that the analyst take an active and conscious part in defining the nature of the problem itself.[24] In actively defining the nature of the problem, analysts must not only impose part of themselves on the problem situation but also exercise reflective judgment and insight.

Creativity in Problem Structuring

The criteria for determining the success of problem structuring are also different from those used to judge the success of problem solving. Successful problem solving requires that analysts obtain relatively precise technical solutions for relatively well-structured problems. By contrast, problem structuring requires that analysts produce creative solutions. Criteria

---

at Berkeley subsequently expanded on the concept, identifying ten characteristics of wicked problems. The ethical nature of the term "wicked" was not discussed. Curiously, at the other end of the continuum, the descriptive term "tame" was simply equated with "satisfying" or "good enough," with no particular ethical connotation.

[21]See, for example, Thomas R. Dye, *Understanding Public Policy*, 3rd ed. (Englewood Cliffs, NJ: Prentice Hall, 1978), pp. 30–31; Richard O. Mason and Ian I. Mitroff, *Creating a Dialectical Social Science* (Dordrecht, The Netherlands: D. Reidel, 1981); and James G. March and Johan P. Olsen, "The New Institutionalism: Organizational Factors in Political Life," *American Political Science Review* 78, 3 (1984): 739–749. Earlier work making many the same points include Herbert A. Simon, "A Behavioral Model of Rational Choice," *The Quarterly Journal of Economics* (1955): 99–118; and David Braybrooke and Charles E. Lindblom. *A Strategy of Decision: Policy Evaluation as a Social Process* (New York: Free Press of Glencoe, 1963).

[22]Rittel and Webber, p. 162.

[23]William N. Dunn, "Methods of the Second Type: Coping with the Wilderness of Conventional Policy Analysis." *Review of Policy Research* 7, 4 (1988): 720–737.

[24]See John R. Hayes, *Cognitive Psychology* (Homewood, IL: Dorsey Press, 1978), pp. 210–213.

for judging creative acts in general are also applicable to creativity in problem structuring. Problem structuring is creative to the extent that:[25]

1. the product of analysis is sufficiently *novel* that most people could not or would not have arrived at the same solution;
2. the process of analysis is sufficiently *unconventional* that it involves the modification or rejection of previously accepted ideas;
3. the process of analysis requires sufficiently *high motivation and persistence* that analysis takes place with high intensity or over long periods of time;
4. the product of analysis is regarded as *valuable* because it provides an appropriate solution for the problem; and
5. the problem as initially posed is so *ambiguous, vague, or ill-defined* that the main challenge is to formulate the problem itself.

Phases of Problem Structuring

Problem structuring has four interrelated phases: *problem search*, *problem delineation*, *problem specification*, and *problem sensing* (Figure 3.3). A first step is the recognition or "felt existence" of a *problem situation,* that is, a diffuse sense of worries, a disbelief, or a surprise. In moving from problem situation, the analyst engages in *problem search.* At this stage, the goal is not the discovery of a single problem (e.g., that of the client or the analyst) but rather the discovery of the problem representations of multiple stakeholders. In this context, practicing analysts usually face a large number of different problem formulations that are distributed throughout the policymaking process. In effect, analysts are faced with a *metaproblem*[26]—a problem-of-problems, so to speak, that involves problem representations so numerous that the number may seem unmanageably huge.[27] Through *problem delineation*, it is possible to define the boundaries of the metaproblem, which is a second-order problem that is the class of all first-order problems, which are its members.[28] Unless these two levels are distinguished, analysts run the risk of formulating the wrong problem by confusing member and class, violating the rule that "whatever involves *all* of a collection must not be one of the collection."[29]

In moving from metaproblem to *substantive problem,* we attempt to *delineate* the problem to its most basic and general terms. For example, we may decide whether the problem is one with roots in economics, sociology, political science, engineering, and so forth. If the substantive problem is conceptualized as an economic one, the analyst

[25]See Alan Newell, J. C. Shaw, and Herbert A. Simon, "The Process of Creative Thinking," in *Contemporary Approaches to Creative Thinking,* ed. H. F. Gruber, G. Terrell, and M. Wertheimer (New York: Atherton Press, 1962), pp. 63–119; see also the excellent monograph by James L. Adams, *Conceptual Blockbusting* (Stanford, CA: Stanford Alumni Association, 1974).
[26]See Yehezkel Dror, *Design for Policy Sciences* (New York: Elsevier, 1971).
[27]This point is made by P. Harmon and D. King, *Expert Systems: Artificial Intelligence in Business* (New York: Wiley, 1985).
[28]Paul Watzlawick, John Weakland, and Richard Fisch, *Change: Principles of Problem Formation and Problem Resolution* (New York: W. W. Norton), p. 6.
[29]Alfred North Whitehead and Bertrand Russell, *Principia Mathematica,* 2nd ed., vol. I (Cambridge: Cambridge University Press, 1910), p. 101. Quoted by Watzlawick et al., p. 6.

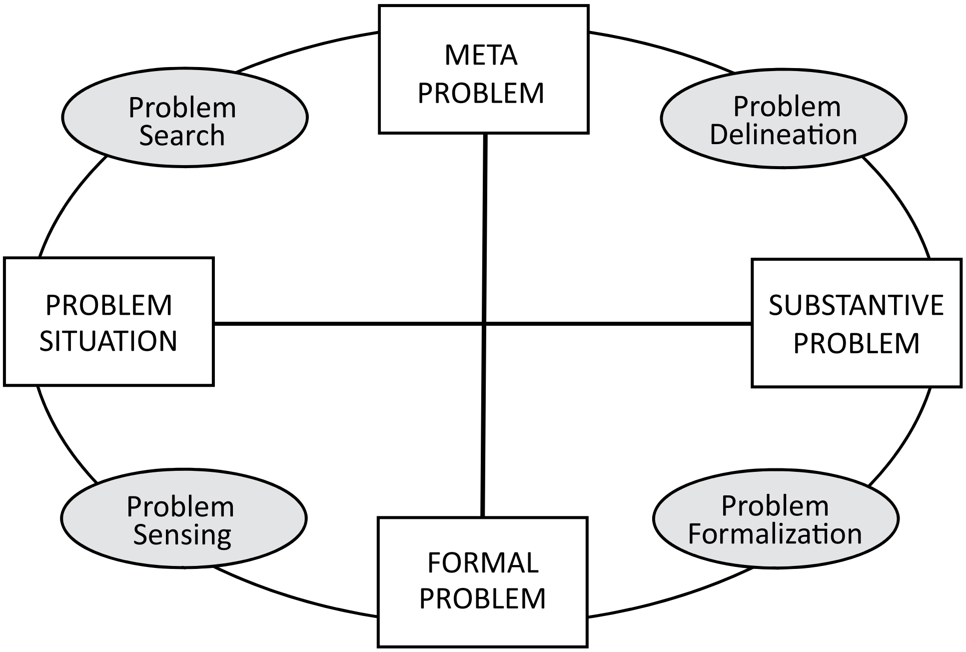


FIGURE 3.3
Phases of Problem Structuring

will treat it in terms of factors related to the production and distribution of goods and services. Alternatively, if the problem is viewed as political or sociological, the analyst will approach it in terms of the distribution of power and influence among competing interest groups, elites, or classes. In many ways, the choice of a conceptual framework is similar to the choice of a worldview, ideology, or popular myth and indicates a commitment to a particular view of reality.[30]

To illustrate the importance of worldviews, ideologies, and popular myths in conceptualizing substantive problems, consider the various ways to define poverty. Poverty may be defined as a consequence of accidents or inevitable conditions of society, of the actions of selfish economic elites, or of imperfections in the poor themselves.[31] These definitions of poverty contain elements of a worldview, myth, or ideology insofar as each involves the selective perception of elements of a problem situation.

Worldviews, ideologies, and myths may be partly true and partly false. In this example, the attribution of poverty to historical accidents may suggest that to alter the distribution of wealth is pointless, but this same myth, which represents a *naturalistic perspective* of social

[30]Ian Mitroff and Ralph H. Kilmann, *Methodological Approaches to Social Science* (San Francisco, CA: Jossey-Bass, 1978). See also Thomas Kuhn, *The Structure of Scientific Revolutions*, 2nd ed. (Chicago, IL: University of Chicago Press, 1971); and Ian G. Barbour, *Myths, Models and Paradigms* (New York: Harper & Row, 1976).
[31]Lowry, *Social Problems,* pp. 19–46.

problems, sensitizes analysts to the relative nature of poverty by pointing to the fact that no known society has wholly resolved it. Similarly, the attribution of poverty to evil or morally corrupt elites may distort their actual motivations. Yet this same *moralistic perspective,* which explains poverty in terms of presumed moral weaknesses, also directs attention to the ways in which the concentration of ownership promotes waste, exploitation, and social irresponsibility. Finally, to attribute poverty to imperfections in the poor themselves results in blaming the victim rather than environing circumstances.[32] Although some poor persons choose to live under conditions that the rest of society defines as "poverty," this perspective may "justifying a perverse form of social action designed to change, not society, as one might expect, but rather society's victim."[33]

Once a substantive problem has been defined, a more detailed and specific *formal problem* may be constructed. The process of moving from substantive to formal problem is carried out through *problem specification*, which involves the development of a formal symbolic or mathematical model of the substantive problem. At this point, difficulties may arise, because the relation between the substantive problem and its formal representation may be tenuous.[34] The specification of a complex problem in simplistic mathematical terms may be premature or inappropriate. Here, the main task is not to obtain the correct mathematical solution but to define the nature of the problem.

Errors of the Third Type (E_{III})

An important issue of problem structuring is how well substantive and formal problems correspond to the problem situation. If a problem situation is a complex system of problems, then a central task is the formulation of substantive and formal problems that adequately represent that complexity. The match between a problem situation and a substantive problem is determined at the problem delineation phase. The substantive problem may be based on assumptions about human nature, time, and the possibilities for change through public action. Equally important, is the fit between the problem situation and the formal problem. The latter is typically specified in the form of a mathematical formula or equation, a process that is subject to all the cautions attending the misspecification of statistical models.

Information about the problem situation is transformed into information about the *metaproblem* through the process of problem search. Analysts who fail to search, or stop searching prematurely, run the risk of inaccurately defining the boundaries of the metaproblem. Important aspects of the metaproblem may be left out, restricting its scope inappropriately. The transformation of information about the metaproblem into information about the substantive problem is accomplished through a process of problem delineation. Here, the risk is selecting from within the boundaries of the metaproblem the wrong substantive problem. In turn, the transformation of the substantive problem into the formal problem

[32]This self-contradictory brand of humanism is known as "blaming the victim." See William Ryan, *Blaming the Victim* (New York: Pantheon Books, 1971).

[33]William Ryan, *Blaming the Victim* (New York: Pantheon Books, 1971), p. 7.

[34]For an example of mis-specification see Ralph E. Strauch, "A Critical Look at Quantitative Methodology," *Policy Analysis* 2 (1976): 121–144.

through a process of problem specification may result in the wrong formal representation, or model, of the substantive problem rather than the right one. Any of the transitions shown in Figure 3.3 may result in a *Type III error*: formulating the wrong problem.[35] A Type III error is different from a Type I statistical error (rejecting the null hypothesis when it is true) and a Type II statistical error (accepting the null hypothesis when it is false). Long ago, Raiffa cautioned that "practitioners all too often make errors of a third kind: solving the wrong problem."[36]

The concept of a Type III error raises issues that are central to policy analysis and science in general. Different phases of problem structuring (Figure 3.3) require different methodological skills. The skills most appropriate for discovering *metaproblems* are primary observational, because the aim is to uncover and aggregate the problem definitions of multiple stakeholders. By contrast, the skills required to transform a metaproblem into a *substantive problem* are conceptual and analytical, much like the skills used in qualitative data analysis. Here, the challenge is to identify a substantive problem by reducing to manageable proportions the multiple definitions of a problem that lie within the boundary of the metaproblem. In turn, the skills required to transform a substantive problem into a formal problem are mathematical, statistical, and logical, including econometric modeling and decision analysis.

POLICY MODELS AND PROBLEM STRUCTURING

Policy models are simplified representations of a problem situation.[37] As such, the development of policy models involves skills in problem structuring. Policy models may be expressed as concepts, diagrams, graphs, or equations and may be used not only to describe, explain, and predict aspects of a problem, but also to solve it by recommending courses of action. Although models are never literal descriptions of a problem, they are useful and even necessary. They simplify systems of problems by reducing and making manageable the complexities encountered by analysts in their work. Policy models also help distinguish essential from nonessential features of a problem, highlight relationships among important factors or variables, and assist in explaining and predicting the consequences of policy choices. Policy models may also play a self-critical role in policy analysis by forcing analysts to make their own assumptions explicit and to challenge conventional wisdom.

The use of policy models is not a matter of choice, because each of us uses models in our work and everyday life. "A mental image is a model. All of our decisions are taken on

[35]See Ian I. Mitroff and Frederick Betz, "Dialectical Decision Theory: A Meta-Theory of Decision-Making," *Management Science,* 19, 1 (1972): 11–24. Kimball defines a Type III error as "the error committed by giving the right answer to the wrong problem." See A. W. Kimball, "Errors of the Third Kind in Statistical Consulting," *Journal of the American Statistical Association* 52 (1957): 133–142.

[36]Howard Raiffa, *Decision Analysis* (Reading, MA: Addison-Wesley, 1968), p. 264.

[37]See Saul I. Gass and Roger L. Sisson, ed., *A Guide to Models in Governmental Planning and Operations* (Washington, DC: Office of Research and Development, Environmental Protection Agency, 1974); and Martin Greenberger, Matthew A. Crenson, and Brian L. Crissey, *Models in the Policy Process* (New York: Russell Sage Foundation, 1976).

the basis of models. The question is not to use or ignore models. The question is only a choice among alternatives.[38]

By simplifying problem situations, models selectively distort reality. Models themselves cannot tell us how to discriminate between essential and nonessential questions; nor can they explain, evaluate, or recommend, because these judgments are external to the model and not part of it. While models help us to undertake these analytic tasks, it is the analyst and not the model who supplies the assumptions necessary to identify features of reality described by a model. Finally, policy models—particularly those expressed in mathematical form—are frequently difficult to communicate to policymakers and other stakeholders who are not versed in the language of mathematical models, although the models can potentially contribute to better decision-making.

Descriptive Models

One type of model, the descriptive model, explains or predicts the causes and consequences of policy choices. Descriptive models are used to monitor the outcomes of policy actions, for example, outcomes such as income, employment, arrest rates, or health care costs included in lists of social and economic indicators published by the Office of Management and the Budget or by Eurostat. Descriptive models are also used to forecast economic performance, energy demand, global warming and other future policy outcomes. For example, the Council of Economic Advisers prepares an annual economic forecast for inclusion in the President's Economic Report.

Normative Models

Another type of model is the *normative model.* Normative models not only attempt to explain or predict, but also to prescribe courses of action for optimizing the attainment of values ranging from income and wealth, education and enlightenment, and crime and justice. Among the many types of normative models used by policy analysts are those that help determine optimum levels of service capacity (queuing models), the optimum timing of service and repairs (replacement models), the optimum volume and timing of orders (inventory models), and the optimum return on public investments (benefit–cost models). Normative decision models take the form: Find the values of the policy variables that will produce the greatest utility, as measured by the value of outcome variables that we wish to change. The values of policy variables and outcome variables are frequently expressed as a monetary quantity such as dollars ($), euros (€), or some other quantity.

A simple and familiar normative model is compound interest. Many persons have used some variation of this model to find the values of a "policy" variable (e.g., putting money in a bank versus a credit union) that will produce the greatest interest income on savings, as measured by the amount of money that one may expect after a given number of years. The model for compound interest is

$$S_n = (1 + r)^n S_0$$

[38]Jay W. Forrester, "Counter-Intuitive Behavior of Social Systems," *Technological Review* 73 (1971): 3.

where S_n is the amount to which savings will accumulate in a given number of years, n, S_0 is the initial savings, and $(1 + r)^n$ is a constant return on investment, 1, plus the rate of interest, r, in a defined time period, n. If we know the interest rates of different savings institutions, and we wish to optimize the return on savings, this simple normative model permits a straightforward choice of the institution offering the highest rate of interest, provided there are no other important considerations (e.g., the security of deposits or special privileges for patrons) that should be taken into account. This normative model of compound interest predicts the accumulation of savings under alternative interest rates, r. This points to a characteristic of all normative models: They not only permit us to estimate past, present, and future values of outcome variables but also allow us to optimize the attainment of some value.

Verbal Models

Models may be expressed in three ways: verbally, symbolically, and procedurally.[39] *Verbal models* are expressed in everyday language, rather than the language of symbolic logic (if $p > q$ and $q > r$, then $p > r$) or mathematics ($\log_{10}10{,}000 = 4$). Verbal models are often parts of policy arguments and expressed as a narrative. Verbal models are relatively easily communicated among experts and laypersons alike, and their costs are low. A potential limitation of verbal models is that the reasons offered for predictions or recommendations may be implicit or hidden, making it difficult to critically examine arguments. Arguments for and against the blockade of the Soviet navy during the Cuban missile crisis are good examples of verbal policy models. President Kennedy's own verbal model of the crisis argued, in effect, that a blockade was the United States' only real option: "Above all, while defending our own vital interests, nuclear powers must avert those confrontations which bring an adversary to a choice of either a humiliating retreat or a nuclear war."[40]

Symbolic Models

Symbolic models use mathematical or logical notation to describe relationships among key variables that are believed to characterize a problem. Predictions or optimal solutions are obtained from symbolic models by employing methods of mathematics, statistics, and logic. Symbolic models are difficult to communicate among laypersons, including policymakers, and even among expert modelers, there are misunderstandings about basic elements of models.[41] The costs of symbolic models are probably not much greater than those of verbal models, considering that we take into account the time and effort expended on public debate, which is the main vehicle for expressing verbal models. A practical limitation of symbolic models is that their results may not be easily interpretable, even among specialists, because the meanings of symbols may not be adequately defined. Symbolic models may

[39]Models may also be expressed physically, as when various materials are used to construct representations of human organs, cities, or machines. The basic limitation of such models is that they cannot represent human action, which involves communication processes, social learning, and choice.
[40]Quoted in Graham T. Allison, "Conceptual Models and the Cuban Missile Crisis," *American Political Science Review* 63, 3 (1969): 698.
[41]See Greenberger and others, *Models in the Policy Process,* pp. 328–336.

improve policy decisions, but only if the assumptions of models are made transparent. "All too frequently what purports to be a model based on theory and evidence is nothing more than a scholar's preconceptions and prejudices cloaked in the guise of scientific rigor and embellished with extensive computer simulations."[42]

Although we have already considered compound interest, a simple symbolic model designed for normative purposes, other symbolic models are descriptive rather than normative. A frequently used descriptive symbolic model is the simple linear equation

$$Y = a + bX$$

where Y is the value of the dependent variable that is to be predicted and X is the value of an independent variable that may be manipulated to achieve Y.[43] The relation between X and Y is known as a linear function, which means that relations between X and Y form a straight line when plotted on a graph (Figure 3.4). In this model, the symbol b, expressed as a coefficient, denotes the amount of change in Y due to a unit change in X. This may be depicted by the slope of a straight line. The steeper the slope, the greater the effect of X on Y. The symbol a, called an *intercept constant*, denotes the point where the straight line intercepts the vertical or Y-axis when X is zero. In Figure 3.4, values of Y are one-half those of X along the broken line (i.e., $Y = 0 + 0.5X$), whereas along the solid line, each change in X produces a 1.0-unit change in Y (i.e., $Y = 0 + 1.0X$). This linear model predicts how much change in the policy variable (X) is needed to produce a value of the outcome variable (Y).

Procedural Models

Procedural models represent dynamic relationships among variables. Predictions and optimal solutions are obtained by searching through and simulating sets of possible relationships, for example, relations among economic growth, energy consumption, and food supplies in future years. Simulation and search procedures are generally performed with the aid of a computer, which is programmed to yield predictions under different sets of assumptions.

Procedural models make use of symbolic modes of expression. The main difference between symbolic and procedural models is that the former use actual data to estimate relationships among policy and outcome variables, whereas procedural models assume, or simulate, such relationships. The costs of procedural models are relatively high, as compared with verbal and symbolic models, largely because of the time required to develop and run computer programs. At the same time, procedural models may be written in reasonably nontechnical language, thus facilitating communication among laypersons. Although the strength of procedural models is that they permit creative simulation and search, it is sometimes difficult to find evidence to justify model assumptions.

A simple form of procedural model is the decision tree, which is created by projecting several possible consequences of policies. Figure 3.5 illustrates a simple decision tree

[42] Gary Fromm, "Policy Decisions and Econometric Models," cited by Greenberger and others, *Models in the Policy Process,* p. 72.

[43] The equation may also be written $Y = b_0 + b_x + e$ in standard notation.

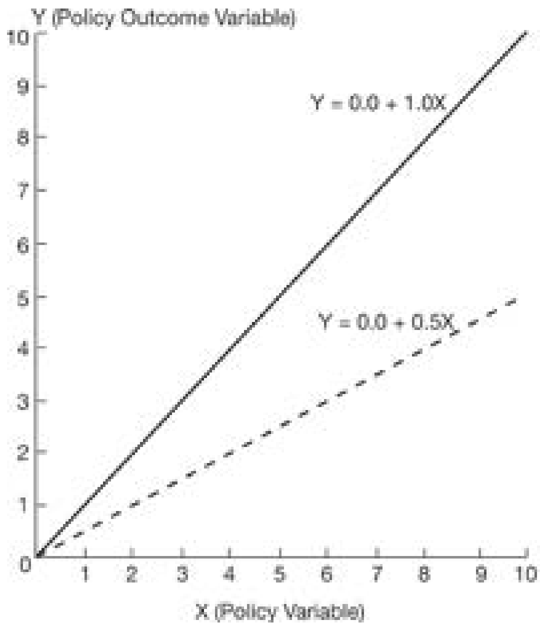


FIGURE 3.4
A Symbolic Model

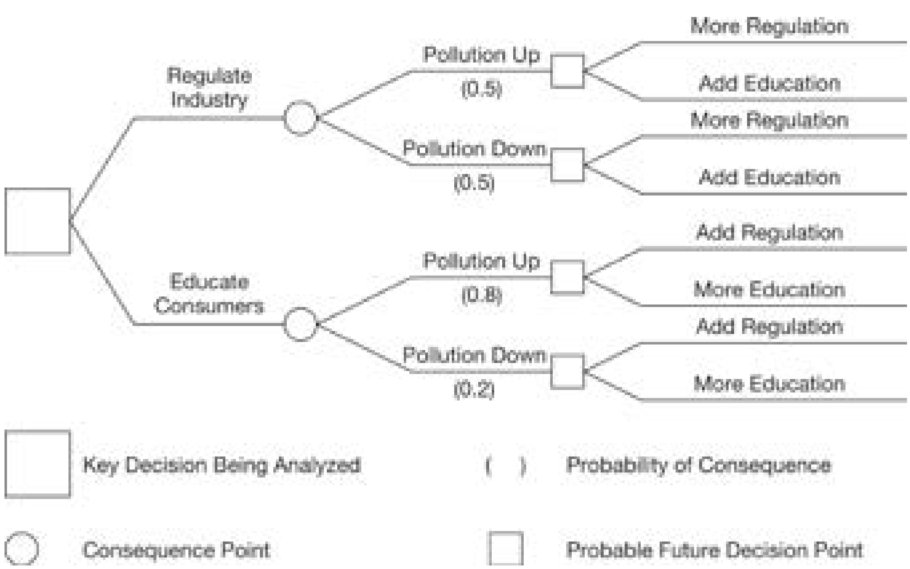


FIGURE 3.5
Simulation or Procedural Model

that estimates the probability that each of several policy alternatives will reduce pollution.[44] Decision trees are useful in comparing subjective estimates of the possible consequences of various policy choices under conditions where there are little or no existing data.

Models as Surrogates and Perspectives

Policy models may be viewed as surrogates or perspectives.[45] A *surrogate model* is assumed to be a substitute or proxy for a substantive problem. Surrogate models are based, consciously or unconsciously, on the assumption that the formal problem is a valid representation of the substantive problem. By contrast, *perspective models* are viewed as one among several possible ways to view a substantive problem. Perspective models assume that the formal problem can never be an entirely valid representation of the substantive problem.

The distinction between surrogate and perspective models is important in public policy analysis, where many of the most important problems are ill structured. The structure of many problems is sufficiently complex that the uncritical use of surrogate models significantly increases the probability of errors of the third type (E_{III})—that is, solving the wrong formulation of a problem when one should have solved the right one. This point may be clarified by considering two illustrations of formal modeling. The first of these is a symbolic model, whereas the second is a verbal model.[46]

Suppose that we have constructed a symbolic model expressed in the form of a linear equation such as that described earlier in Figure 3.4. Using the equation $Y = a + b_X$ (later we will use standard notation to write this equation as $Y = b_0 + b_x + e$), we may begin by plotting actual values of X and Y on the scatterplot, as shown in Figure 3.6. Our implicit perspective is that the observed values of X and Y constitute a surrogate model of a causal relationship, where the policy variable, X, is believed to be the cause of the outcome variable, Y. We might erroneously interpret the results as a surrogate model, one that confirms the supposed causal structure of the substantive problem. We may interpret the slope of the line as a measure of the effects of the depth of the reservoir, X, on Y, the rainfall rate. In addition, the correlation between observed values of Y (data points) and those predicted by the equation (those lying on the straight line) will be taken as a confirmation of the accuracy of the prediction. The probable conclusion is that a change in the value of the policy variable will result in a corresponding change in the value of the outcome variable. This is a Type III error.[47]

The point of this illustration is that the formal symbolic model provides no guidance in answering the question of whether X causes Y, or Y causes X. If X happens to be unemployment and Y poverty, there may be a case for the predicted relation, provided we provide plausible reasons for believing that unemployment is causally relevant to poverty. Yet

[44]See Gass and Sisson, *A Guide to Models in Governmental Planning and Operations*, pp. 26–27.

[45]Strauch, "A Critical Look at Quantitative Methodology," pp. 136–144.

[46]These illustrations are adapted from Strauch, "A Critical Look at Quantitative Methodology," pp. 131–133; and Watzlawick, Weakland, and Fisch, *Change.*

[47]Of course if we were simply correlating the depth of reservoir with rainfall we would reach a correct answer. It is when we manipulate the independent (policy) variable to produce rainfall that we encounter difficulties not unlike those of a failed shaman.

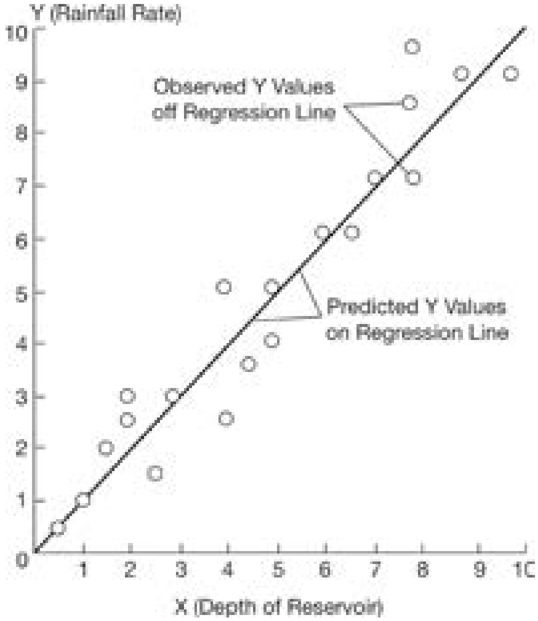


FIGURE 3.6
Assumed Effects of *X* on *Y* from One Perspective

this causal information is not contained in the symbolic model; it comes from *outside* the model. The observed relation between *X* and *Y* might just as easily be interpreted the other way, as evidence that poverty affects unemployment, for example, by presuming that poverty depresses the desire to seek employment.

The lesson of this illustration is that the conceptualization of substantive problems governs their specification as formal symbolic models. Formal models of substantive problems may conceal the fact that the formal model is a perspective of "reality" and not a surrogate. To clarify the point, consider the formal symbolic model previously discussed. Suppose that *X* is the mean annual depth of a reservoir and that *Y* is the annual rainfall in the surrounding area. Because policies governing reservoir depth can be changed, reservoir depth is a policy variable subject to manipulation. If we wish to control annual rainfall, common sense would suggest that the relation between rainfall and reservoir depth is a causal relation. However, the conclusion that we can decrease rainfall by draining water more rapidly from the reservoir is invalid and nonsensical. As RAND analyst Ralph Strauch has emphasized, "This is because the causal relationship assumed in the analysis—reservoir depth causes rainfall—runs counter to our common sense understanding that rainfall determines reservoir depth."[48]

Consider a second illustration of the difficulties that arise when we confuse surrogate and perspective models. This time we will use an illustration that makes the same point about verbal models, that is, models that are expressed in everyday language. Suppose that we are confronted with the following problem situation: The director of the state department of transportation has requested that a study be undertaken to

[48]Strauch, "A Critical Look at Quantitative Methodology," p. 132.

recommend the least costly way to connect nine key transportation points in the central region of the state. As the agency's policy analyst you are directed to show how all nine points may be connected by four sections of highway. You are also told that these four sections of highway must be straight (no curves will be permitted) and that each new section must begin at the point where the last section stopped (the construction team will not be permitted to retrace its steps). You are then shown a map of the region (Figure 3.7) and asked to make a recommendation that will solve the director's problem.

Unless you were already familiar with this classic conceptual problem (called simply the "nine-dot problem"), it is very unlikely that you were able to solve it. Few people manage to find the solution by themselves, not because the problem is technically complicated (in fact it is simple) but because people almost always commit an error of the third type (E_{III}), solving the wrong formulation of a problem when they should have solved the right one. This is because most people approach the *problem situation* with implicit assumptions that make it impossible to solve the problem. These assumptions are introduced by analysts themselves; they are *not* part of the problem situation. In other words, analysts themselves create a substantive problem that cannot be solved.

The solution for the nine-dot problem is presented in Figure 3.8. The solution appears surprisingly simple, novel, and unconventional; that is, it has several of the key characteristics of creativity discussed earlier. In the process of analysis, it becomes suddenly clear

FIGURE 3.7
Map of Transportation Points in Central Region

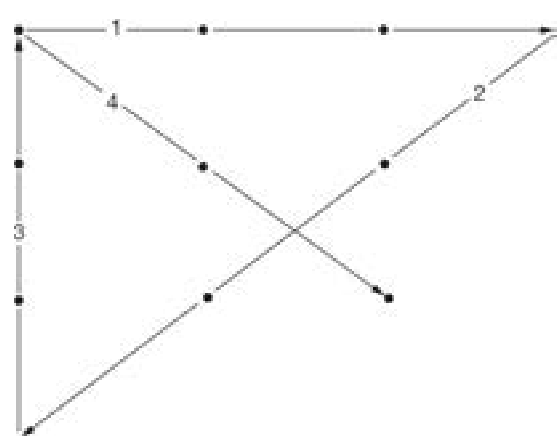


FIGURE 3.8
Solution for Nine-Dot Problem

that we have been solving the wrong problem. So long as we assume that the solution must be found within the boundaries set by our verbal model—that is, a rectangle composed of nine dots—then a solution is not possible. Yet this condition is imposed not by the formal verbal model but by the implicit assumption of "squareness" that unconsciously shapes our definition of the substantive problem. Imagine what would have occurred if we had transformed the verbal model into a symbolic one, for example, by using plane geometry to derive precise quantitative estimates of the distance between points. This would not only have led us further and further away from a solution; it would also have created an aura of scientific precision that would lend authority to the conclusion that the problem is "insoluble." This illustration helps convey a simple but important point about the use of formal models: formal models cannot themselves tell us whether we are solving the wrong formulation of a problem. *The map is not the territory! The menu is not the meal!*[49]

METHODS OF PROBLEM STRUCTURING

Problem structuring is the process of generating and testing alternative formulations of substantive problems. As we saw in Figure 3.3, problem structuring has four interrelated phases: problem sensing, problem search, problem delimitation, and problem specification. A number of methods and techniques are helpful in carrying out problem-structuring activities in each phase. These methods, along with their respective aims, procedures, sources of knowledge, and criteria of performance, are displayed in Table 3.2.

Boundary Analysis

An important task of problem structuring is estimating whether the system of individual problem formulations that we have called a *metaproblem* is relatively complete. This task is similar to the situation of the homesteaders described by mathematician Morris Kline in his essays on the myth of certainty in mathematics.[50] The homesteaders, while clearing their land, are aware that enemies lurk in the wilderness that lies just beyond the clearing. To increase their security, the homesteaders clear a larger and larger area but never feel completely safe. Frequently, they must decide whether to clear more land or attend to their crops and domesticated animals within the boundaries of the clearing. They do their best to push back the wilderness but know that the enemies lurking beyond the clearing could surprise and overcome them. They hope that they will not choose to tend the crops and livestock when, instead, they should have been clearing more land.

The analogy of the homesteaders points to a key problem of problem structuring. We are rarely faced with a single, well-defined problem; rather, we are faced with multiple problems that are defined in different ways by stakeholders. Analysts are analogous to

[49]These admonitions are those of Gregory Bateson, *Steps to an Ecology of Mind: Collected Essays in Anthropology, Psychiatry, Evolution, and Epistemology* (Chicago, IL: University of Chicago Press, 1972).

[50]Morris Kline, *Mathematics: The Loss of Certainty* (New York: Oxford University Press, 1980). For essays on uncertainty and creativity in modern mathematics, see Michael Guillen, *Bridges to Infinity* (Ithaca, NY: Cornell University Press, 1988).

TABLE 3.2

Characteristics of Methods of Problem Structuring

Method	Aim	Procedures	Source of Knowledge	Performance Criterion
Boundary Analysis	Estimate boundaries of metaproblem	Saturation sampling and problem elicitation	Multiple Stakeholders	Correctness-in-limit
Classification Analysis	Clarification of concepts	Logical division and classification of concepts	Single Analyst	Logical Consistency
Hierarchy Analysis	Identify possible, plausible, and actionable causes	Logical division and classification of causes	Single Analyst	Logical Consistency
Synectics	Recognition of similarities among problems	Construction of personal, direct, symbolic, and fantasy analogies	Group Members	Plausibility
Brainstorming	Generation of ideas, goals, and strategies	Idea generation and evaluation	Group Members	Consensus
Multiple Perspective Analysis	Generation of insight	Use of technical, organizational, and personal perspectives	Group Members	Multiplicity of Perspectives
Assumption Analysis	Creative synthesis of conflicting assumptions	Stakeholder identification, assumption surfacing, challenging, and synthesis	Group Members	Dialectical Productivity
Argument Mapping	Critical thinking	Plausibility analysis of arguments	Group Members	Plausibility

homesteaders who, working without known boundaries, have no "stop rule."[51] They are modern counterparts of Diogenes, engaged in "a never-ending discourse with reality, to discover yet more facets, more dimensions of action, more opportunities for improvement."[52]

To make effective use of the methods of problem structuring described in this chapter, it is important to conduct a *boundary analysis*. Methods of problem structuring just discussed, along with related methods that presuppose that the problem has already been structured,[53] do not themselves provide a way to know whether a set of problem formulations is relatively complete.

The relative completeness of a set of problem formulations may be estimated by means of a three-step process.[54]

1. *Saturation sampling.* The first step is drawing a saturation (or snowball) sample of stakeholders. Stakeholders in this initial set may be contacted, face-to-face or by telephone or computer, and asked to name two additional stakeholders, one of whom agrees most and one least with the arguments. The process is continued until no *new* stakeholders are named. There is little or no sampling variance, because all or most members of the working universe of policy-relevant stakeholders in a specific area (e.g., health care reform) are contacted.[55]
2. *Elicitation of problem representations.* In the second step analysts elicit alternative problem representations from stakeholders. These are the problem representations that Heclo describes as "ideas, basic paradigms, dominant metaphors, standard operating procedures, or whatever else we choose to call the systems of interpretation by which we attach meaning to events."[56] The information required to document these problem representations may be obtained from face-to-face interviews or, given the time constraints facing most analysts, from telephone and computer communications and available documents.
3. *Boundary estimation.* The third step is to estimate the boundaries of the metaproblem. A cumulative frequency distribution displays stakeholders on the horizontal axis and the number of *new* problem elements, which may be assumptions, ideas, objectives, or policies, are plotted on the vertical axis (Figure 3.9). As the new (i.e., nonduplicative) elements of each stakeholder are plotted, the slope of the curve

[51]The "stop rule" originated with Rittel and Webber, "Dilemmas in a General Theory of Planning," *Policy Sciences* 4 (1973): 155–169.

[52]Dery, *Problem Definition in Policy Analysis*, pp. 6–7.

[53]These other methods include the analytic hierarchy process, interpretive structural modeling, policy capturing, and Q-methodology. See, respectively, Thomas L. Saaty, *The Analytic Hierarchy Process* (New York: McGraw-Hill, 1980); John N. Warfield, *Societal Systems: Planning, Policy, and Complexity* (New York: Wiley, 1976); Kenneth R. Hammond, *Judgment and Decision in Public Policy Formation* (Boulder, CO: Westview Press, 1977); and Stephen R. Brown, *Political Subjectivity: Applications of Q-Methodology in Political Science* (New Haven, CT: Yale University Press, 1980). Computer software is available for applications of these methods.

[54]Dunn, "Methods of the Second Type: Coping with the Wilderness of Conventional Policy Analysis."

[55]On these points as they apply to sociometric and saturation sampling in general, see Seymour Sudman, *Applied Sampling* (New York: Academic Press, 1976).

[56]Hugh Heclo, "Policy Dynamics," in *The Dynamics of Public Policy,* ed. Richard Rose (Beverly Hills, CA: Sage Publications, 1976), pp. 253–254.

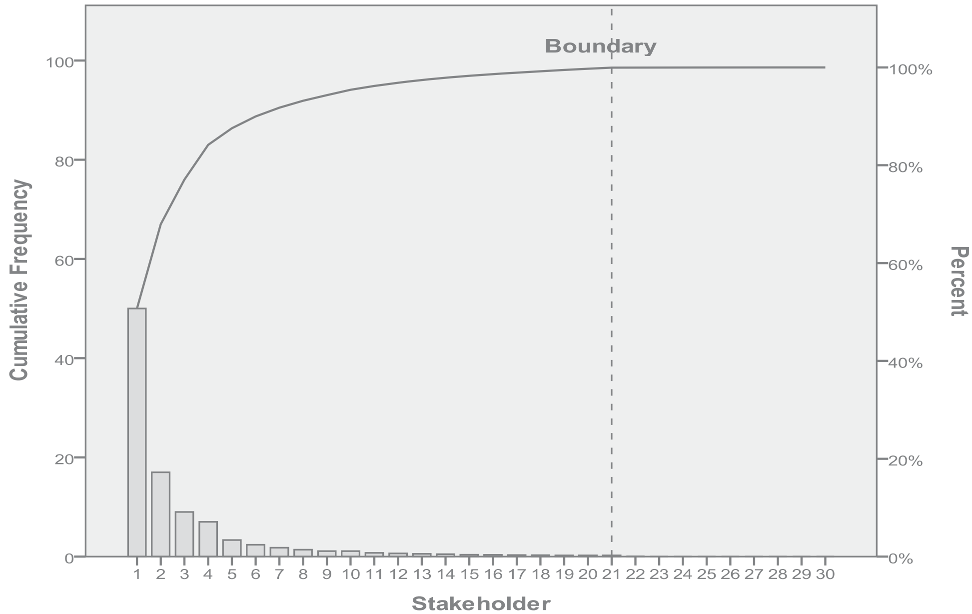


FIGURE 3.9
Pareto Chart—Estimated Boundary of Problem

displays different rates of change. An initial rapid rate of change is followed by slow change and eventually stagnation. This is the point at which the curve becomes flat. After this point, the collection of additional information about elements of the problem is unlikely to improve the completeness of the set, because the limit of the boundary of the metaproblem has been estimated.

These estimation procedures satisfy requirements for sound inductive estimates in general: character, coordination, cost-effectiveness, and correctness-in-the-limit.[57] Applications of similar procedures in other areas—for example, estimates of boundaries of scientific literature, library holdings, languages, literary works, consumer preferences—suggest lawful regularities in the growth of knowledge systems.[58] Boundary analysis, like other policy-analytic procedures, yields results that are plausible and not certain. When used in conjunction with other problem-structuring methods, boundary estimation procedures reduce the likelihood of Type III errors in policy analysis.

---

[57]See Nicholas Rescher, *Induction* (Pittsburgh, PA: University of Pittsburgh Press, 1980), pp. 24–26. For detailed arguments relating these requirements to problem structuring, see Dunn, "Methods of the Second Type."
[58]For a brief but provocative essay on these regularities, see Herbert A. Simon, "The Sizes of Things," in *Statistics: A Guide to the Unknown,* ed. Judith M. Tanur and others (San Francisco, CA: Holden-Day, 1972), pp. 195–202.

Classification Analysis

Classification analysis is a technique for clarifying concepts used to define and classify problem situations.[59] In sensing a problem situation, analysts must classify their experiences. Even simple descriptions of problem situations are based on the classification of experience through inductive reasoning, a process where general (abstract) concepts, such as poverty, crime, and pollution, are formed by experiencing particular (concrete) objects or situations. When we classify a problem situation in one way, we foreclose opportunities to classify it in another way, as the nine-dot problem illustrates.

Classification analysis is based on logical division and logical classification. When we select a class of elements and break it down into its component parts, the process is called *logical division;* the reverse process, which involves the combination of elements into larger groups or classes, is called *logical classification.* The basis of a classification depends on our purpose, which in turn depends on substantive knowledge about a problem.

Consider poverty in the United States. As a percent of the total population, poverty increased from 11.2 percent in 2000 to 14.9 percent and 14.6 percent in 2010 and 2014, respectively. The percent of persons below the poverty level ($10,890 per person in 2011) may be broken down into three age categories: less than 18 years, 18–64 years, and more than 64 years. If we use logical division among the dates in row 1 of Table 3.3, we will reach the conclusion that the problem of poverty in the United States is about the same in 2000 as it was in 1970, when President Lyndon Johnson's War on Poverty was implemented. When the process of logical division is carried one step further, and persons below the poverty level are divided into the three classes on the basis of age, we reach a different conclusion. Now we conclude that poverty has become slightly worse for adults aged 18–64 and for children under 18 years of age, compared with persons aged over 64, where poverty declined. The welfare reforms implemented in 1996 under President Clinton appear to have benefitted all groups. It also appears that poverty is not being alleviated by the private enterprise system, because the percent of persons below the poverty line has increased for all groups except elders, only 5–6 percent of whom earn employment income and almost all have the benefits of government support (Table 3.3). The recession of 2008 also had an effect on poverty. Compared with 2000, all groups were worse off except elders, although there was a slight overall improvement in 2010–2014 under President Obama.

There is no way to know with certainty if the categories of a classification system are the right ones. But there are several rules that help ensure that a classification system is both relevant to a problem situation and logically consistent:

1. *Substantive relevance.* The basis of a classification should be developed according to the analyst's purpose and the nature of the problem situation. This deceptively simple rule, means that classes and subclasses should conform as closely as possible to the "realities" of the problem situation. However, what we know about a situation is partly a function of the concepts we use to experience it. For this reason, there are no absolute guidelines that tell us when we have perceived a problem correctly. Poverty, for example, may be classified as a problem of inadequate income, unemployment, cultural deprivation, or psychological motivation—it may be all of these and more.

[59]John O'Shaughnessy, *Inquiry and Decision* (New York: Harper & Row, 1973) pp. 22–30.

TABLE 3.3

Percent of Persons Living below the Poverty Level by Age Group, 1970–2014

Age Groups	1970	1980	1990	2000	2010	2014
All	12.5	12.9	13.5	11.2	14.9	14.6
Percent <18	15.1	18.3	20.6	16.2	22.0	21.1
Percent 18–64	9.0	10.1	10.7	9.6	13.8	13.5
Percent >64	24.6	15.7	12.2	9.9	8.9	10.0

Note: In 2011, the U.S. Bureau of the Census defined poverty as an individual earning less than $10,890, with an added $3,820 per household member ($22,350 for a family of four). Data exclude government transfers in the form of cash payments (Social Security, Public Assistance), nutrition (Food Stamps), housing, health (Medicaid, Medicare), social services (OEO), employment and manpower, and education.

Source: U.S. Census Bureau, www.census.gov/topics/income-poverty/poverty/guidance/poverty-measures.html.

2. *Exhaustiveness.* Categories in a classification system should be exhaustive. This means that all subjects or situations of interest must fit into one or another of the various categories. If we discover that some have additional income, either from their own efforts or government transfers, a new category might be created.
3. *Disjointness.* Categories must be mutually exclusive. Each subject or situation must be assigned to one and only one category or subcategory. In classifying families, for example, they must fall into one or the other of the two main subcategories (income above and below the poverty line), which means that no family can be "double-counted."
4. *Consistency.* Each category and subcategory should be based on a single classification principle. A violation of this rule leads to overlapping subclasses and is known as the fallacy of cross-division. For example, we commit the fallacy of cross-division if we classify families according to whether they are above the poverty line or receive welfare payments, because many families fall into both categories. This rule is actually an extension of rules of exhaustiveness and disjointness.
5. *Hierarchical distinctiveness.* The meaning of levels in a classification system (categories, subcategories, sub-subcategories) must be carefully distinguished. The rule of hierarchical distinctiveness, which is a guideline for interpreting classification systems, is derived from the principle discussed previously: Whatever involves *all* of a collection must not be one of the collection. Humankind is the class of all individuals; but it is not itself an individual. Similarly, poverty is a characteristic of 45 million persons, but they cannot be understood in terms of the behavior of one person multiplied 45 million times. Thus, a population of 45 million persons is not just quantitatively different from a single family; it is also qualitatively different because it involves an aggregate or a system.

The rule of hierarchical distinctiveness is central to problem structuring. In structuring problems, it frequently happens that analysts ignore the distinction between member and class and forget that a class cannot be a member of itself. This point can be illustrated by returning to the nine-dot problem (Figure 3.7). When a person first attempts to solve this problem, the assumption is that the dots compose a square or a rectangle and that the solution must be found within it. This condition is self-imposed; it is not part of the problem as given by the instructions. The failure to solve the nine-dot problem occurs because we ourselves create the problem. It does not matter which combination of four lines we try. We always fail to connect one dot. We can try all possibilities, but none work, because they are confined to the class defined as a square. We eventually learn that "the solution is a second-order change [i.e., one that involves *all* of the class] which consists in leaving the field and which cannot be contained within itself because . . . it involves all of a collection and cannot, therefore, be part of it."[60]

A useful approach to classification analysis is *set thinking*.[61] Set thinking involves thinking about the relations of sets to each other and to subsets, where a *set* is defined as a clearly delimited collection of objects or elements. Sets and subsets are the equivalents of classes and subclasses in a classification system and are expressed visually with the aid of Venn diagrams.[62] In the Venn diagram in Figure 3.10, the rectangle is used to represent all families in the United States. In set language, it is known as the *universal set* (*U*). Two of its component sets, shown as circles *A* and *B*, represent families above and below the poverty line. If we apply rules of exhaustiveness, disjointedness, and consistency, all families will be divided into one or the other of the sets, *A* and *B,* so that the two sets reflect distinct levels of poverty that do not overlap. In set language, the *union* of *A* and *B* is equal to the *universe* (*U*) of all families. The symbol for union is   , read as "union" or "cup."

Two sets may intersect to form a subset, so that the properties of the original two sets overlap, as shown in Figure 3.11. For example, the intersection (*D*) of nonpoverty (*A*) and poverty (*B*) families might be used to illustrate that some families in each group receive government transfer payments. In set language, the intersection of *A* and *B* is equal to *D,* which is expressed symbolically as A   B = D and read "*A* intersect (or cap) *B* equals *D.*" Union and intersection are the two most important set operations and may be used to construct classification schemes (Figure 3.12) and cross breaks (Figure 3.13). Cross breaks are a form of logical division used to organize data in tables.

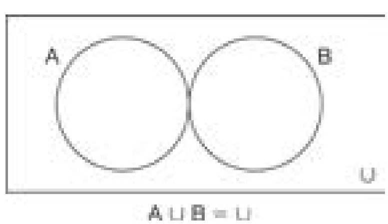


FIGURE 3.10
Set Union

[60]Watzlawick et al., *Change,* p. 25. Gregory Bateson, *Steps to an Ecology of Mind* (New York: Ballantine Books, 1972), cautions that we must not confuse the map with the territory, the menu with the meal.
[61]Set theory, created by the German mathematician Georg Canter (1874–1897), is the mathematical theory of collections of aggregates of entities.
[62]Venn diagrams, used extensively to illustrate set problems, are named after the English logician, John Venn (1834–1923).

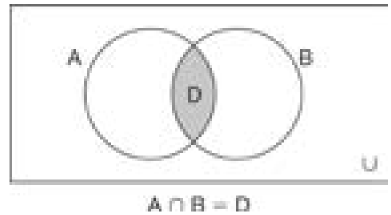


FIGURE 3.11
Set Intersection

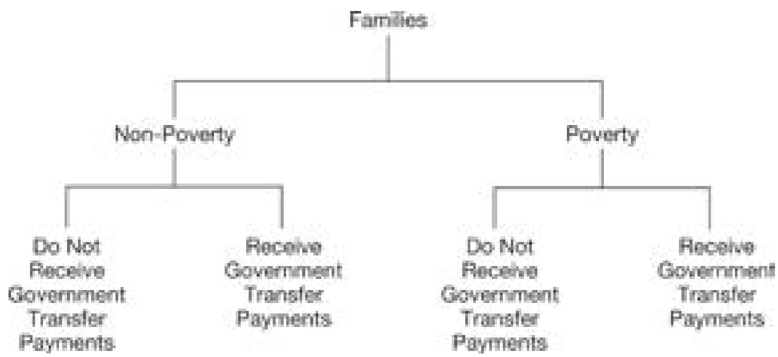


FIGURE 3.12
Classification Scheme

FIGURE 3.13
Cross break

	A_1	A_2
B_1	B_1A_1	B_1A_2
B_2	B_2A_1	B_2A_2

Venn diagrams, classification schemes, and cross breaks are important tools for structuring problems. Classification analysis is usually (but not necessarily) employed by individuals, rather than groups, and logical consistency is the criterion for assessing how well a problem has been conceptualized. While logical consistency is important, there is no way to know with complete confidence that the substantive basis of a category is the right one. While classification analysis improves the clarity of concepts and their relationships, it guarantees only that concepts have approximate substantive relevance.

Hierarchy Analysis

Hierarchy analysis is a technique for identifying possible causes of a problem.[63] Formal logic and many social science theories provide little guidance in identifying possible causes. There is no acceptable way to formally deduce causes from effects, or effects from causes, and social science theories are frequently so general as to be of little help. In order to identify

[63]See O'Shaughnessy, *Inquiry and Decision,* pp. 69–80.

possible causes, it is useful to have a conceptual framework that identifies the many causes that may be operating in a given situation.

Hierarchy analysis helps identify three kinds of causes: possible causes, plausible causes, and actionable causes. *Possible causes* are events or actions that, however remote, may contribute to the occurrence of a given problem situation. For example, unwillingness to work, unemployment, and the distribution of power and wealth among elites may all be taken as possible causes of poverty. By contrast, *plausible causes* are those that, on the basis of research or direct experience, are believed to be an important influence on the occurrence of a situation judged to be problematic. In the preceding example, unwillingness to work is unlikely to be regarded as a plausible cause of poverty, at least among experienced observers, whereas unemployment is. Finally, the distribution of power and wealth among elites is unlikely to be viewed as an *actionable cause*—that is, one that is subject to manipulation by policymakers—because no policy intended to resolve problems of poverty can alter the social structure of an entire society. In this example unemployment is both a plausible and actionable cause of poverty.

Nagel and Neef provide a good example of the potential uses of hierarchy analysis. Many observers have been ready to accept the explanation that the major cause of overcrowded jails is the large number of persons arrested and detained in jail while awaiting trial. For this reason, a policy of pretrial release—that is, a policy that provides for the release of a certain number of those arrested (usually for less serious offenses) prior to a formal trial—has been favored by reformers.

The difficulty with this policy, as well as the causal explanation on which it is based, is that it overlooks plea bargaining as one among several plausible causes of jail overcrowding. In plea bargaining, which is widely practiced in the U.S. judicial system, a defendant agrees to plead guilty in return for a prosecutor's agreement to reduce or drop the charge. When plea bargaining is taken into account along with pretrial release, the following consequences may occur:

- If the percentage of defendants released prior to trial goes up, then the percentage of successful plea bargains will probably go down.
- If guilty pleas go down as a result of increased pretrial release, then the number of trials will probably go up.
- If the number of trials increases, then the delay in going to trial will also increase, unless the system increases the number of prosecutors, judges, and public defenders.
- If the delay in going to trial increases for cases in general, including defendants in jail, then the jail population may increase. Its size depends on the length of time defendants are kept in jail and on the quantity of defendants.
- Any decrease in the jail population caused by pretrial release may be more than offset by the increased delay and length of pretrial detention, which in turn is caused by increased pretrial release, the consequent reduction in guilty pleas and the increase in trials.[64]

This example not only illustrates the potentially creative role of hierarchy analysis in problem structuring. It also shows how hierarchy analysis can help discover possible

[64]Stuart S. Nagel and Marian G. Neef, "Two Examples from the Legal Process," *Policy Analysis* 2, 2 (1976): 356–357.

unanticipated consequences of policies. For example, what could be more obvious than that pretrial release will result in a reduction of the jail population? The answer depends on having a satisfactory understanding of the plausible causes that contribute to the original problem situation. Figure 3.14 provides another illustration of hierarchy analysis applied to possible, plausible, and actionable causes of fires.

The rules for conducting a hierarchy analysis are the same as those used for classificational analysis: substantive relevance, exhaustiveness, disjointness, consistency, and hierarchical distinctiveness. Similarly, procedures of logical division and classification also apply to both types of analysis. The main difference between classificational analysis and hierarchy analysis is that the former involves the division and classification of concepts in general, whereas hierarchy analysis uses concepts of possible, plausible, and actionable causes. Nevertheless, both forms of analysis are used by individuals who employ logical consistency as the primary criterion for assessing how well a problem has been conceptualized. However,

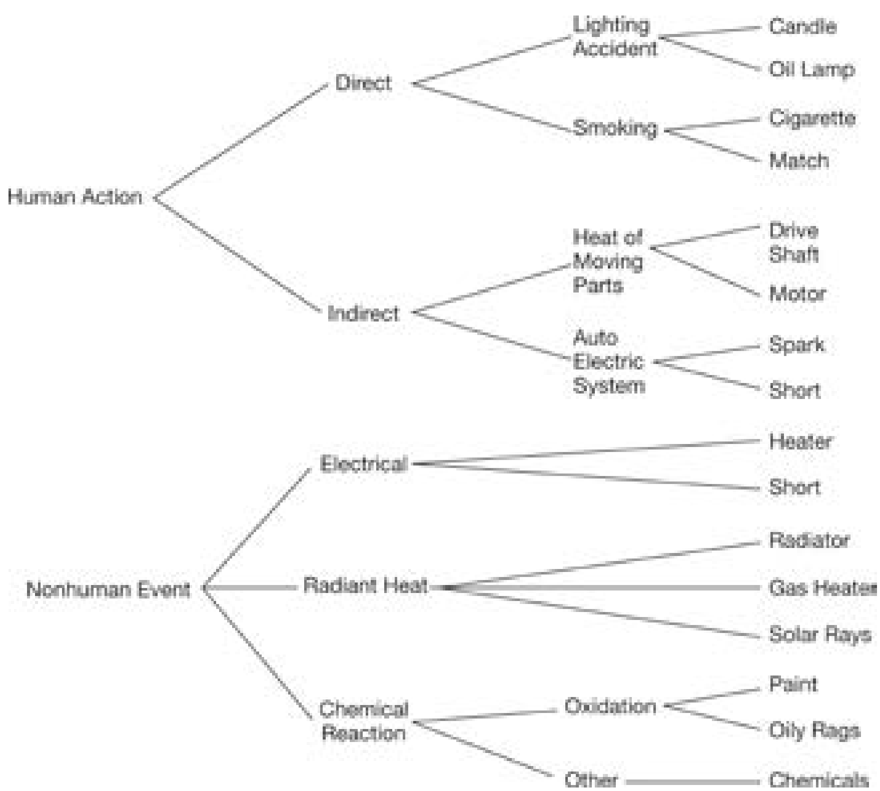


FIGURE 3.14
Hierarchy Analysis of the Causes of Fires

Source: O'Shaughnessy (1973): 76.

neither guarantees that the correct substantive basis for concepts will be found. Hierarchy analysis may foreclose opportunities to generate alternative causal explanations if it relies on the individual analyst, rather than on groups, as a source of knowledge.

Synectics

Synectics is a method designed to promote the recognition of analogous problems.[65] Synectics, which refers broadly to the investigation of similarities, helps analysts make creative use of analogies in structuring problems. Many studies show that people fail to recognize that what appears to be a new problem is really an old one in disguise and that old problems may contain potential solutions for problems that appear to be new. Synectics is based on the assumption that an awareness of identical or similar relationships among problems increases the problem-solving capacities of analysts.

In structuring policy problems, four types of analogies may be identified:

1. *Personal analogies.* In constructing personal analogies, analysts attempt to imagine themselves experiencing a problem situation in the same way as does someone else, for example, a policymaker or client. Personal analogies are often important in uncovering political dimensions of a problem situation, for "unless we are willing and able to think 'politically'—if only as a matter of role playing—we will not be able to enter the phenomenological world of the policymaker and understand the policy process."[66]
2. *Direct analogies.* In making direct analogies, the analyst searches for similar relationships among two or more problem situations. In structuring problems of drug addiction, for example, analysts may construct direct analogies from experiences with the control of contagious diseases.[67]
3. *Symbolic analogies.* In making symbolic analogies, the analyst attempts to discover similar relationships between a given problem situation and some symbolic process. For example, symbolic analogies are often drawn between servomechanisms of various kinds (thermostats, automatic pilots) and policy processes. In each case, analogous processes of adaptation are viewed as consequences of continuous feedback from the environment.[68]
4. *Fantasy analogies.* In making fantasy analogies, analysts are completely free to explore similarities between a problem situation and some imaginary state of affairs. Defense policy analysts, for example, have used fantasy analogies to structure problems of defense against nuclear attack.[69]

---

[65]See W. J. Gordon, *Synectics* (New York: Harper & Row, 1961); and Hayes, *Cognitive Psychology,* pp. 72, 241.
[66]Raymond A. Bauer, "The Study of Policy Formation: An Introduction," in *The Study of Policy Formation,* ed. R. A. Bauer and K. J. Gergen (New York: Free Press, 1968), p. 4.
[67]Mark H. Moore, "Anatomy of the Heroin Problem: An Exercise in Problem Definition," *Policy Analysis* 2, 4 (1976): 639–662.
[68]David Easton, *A Framework for Political Analysis* (Englewood Cliffs, NJ: Prentice Hall, 1965).
[69]Herman Kahn, *On Thermonuclear War* (Princeton, NJ: Princeton University Press, 1960). For a critique, see Philip Green, *Deadly Logic: The Theory of Nuclear Deterrence* (Columbus: Ohio State University Press, 1966).

Synectics relies on individuals as well as groups to create analogies. The main criterion for assessing how well a problem has been conceptualized is the plausibility of comparisons, that is, the degree to which a given problem situation is similar to others taken as analogies.

Brainstorming

Brainstorming is a method for generating ideas, goals, and strategies that may help conceptualize problem situations. Originally designed by Alex Osborn as a means to enhance creativity, brainstorming may be used to generate a large number of suggestions about potential solutions for problems.[70] Brainstorming involves simple procedures:

1. Brainstorming groups should be composed in accordance with the nature of the problem situation being investigated. This usually means the selection of persons who are particularly knowledgeable about the given situation.
2. Processes of idea generation and idea evaluation should be kept strictly apart, because intense group discussion may be inhibited by premature criticism and debate.
3. The atmosphere of brainstorming activities should be kept as open and permissive as possible in the idea-generating phase.
4. The idea-evaluating phase should begin only after all ideas generated in the first phase have been exhausted.
5. At the end of the idea-evaluating phase, the group should prioritize ideas and incorporate them in a proposal that contains a conceptualization of the problem and its potential solutions.

Brainstorming is a versatile procedure that may involve relatively structured or unstructured activities, depending on the aims and the practical constraints of the situation. Relatively unstructured brainstorming activities occur frequently in government agencies and public and private "think tanks." Here discussions of policy problems are informal and largely spontaneous, involving the interaction of generalists and specialists from several scientific disciplines or fields.[71]

Brainstorming-like activities may also be relatively structured, with various procedures used to coordinate or focus group discussions. These devices may include the use of Harold Lasswell's *continuous decision seminars* that, seeking to avoid the restrictive atmosphere of conventional committees, involve a team of highly motivated experts who meet with high frequency over a number of years.[72]

---

[70] Alex F. Osborn, *Your Creative Power* (New York: Charles Scribner, 1948).

[71] See Edgar S. Quade, *Analysis for Public Decisions* (New York: American Elsevier Publishing, 1975), pp. 186–188; and Olaf Helmer and Nicholas Rescher, *On the Epistemology of the Inexact Sciences* (Santa Monica, CA: RAND Corporation, February 1960).

[72] Harold D. Lasswell, "Technique of Decision Seminars," *Midwest Journal of Political Science* 4, 2 (1960): 213–226; and Lasswell, *The Future of Political Science* (New York: Atherton Press, 1963).

Another set of procedures for coordinating and focusing brainstorming activities is the construction of *scenarios,* which are outlines of hypothetical future events that may alter some problem situation. Scenario writing, which has been used to explore potential military and political crises, involves the constructive use of imagination to describe some aspect of a future situation. There are two major types of scenarios: operations-analytical and free-form. In constructing a free-form scenario, the analyst is "an iconoclast, a model breaker, a questioner of assumptions, and—in rare instances—a fashioner of new criteria."[73] By contrast, an operations-analytical scenario has limited aims. Instead of using fantasies or utopias, the operations-analytical scenario "starts with the present state of the world and shows how, step by step, a future state might evolve in a plausible fashion out of the present one."[74]

A good example of relatively structured brainstorming is the Year 2000 Planning Program, a two-and-a-half year project carried out in the U.S. Bureau of the Census.[75] In this project, 120 self-selected participants from all levels and branches of the Bureau, from secretaries to division heads and the director, were asked to think as freely as possible about the future. In effect, they constructed free-form scenarios, which to them indicated what the bureau should be like in the year 2000. Participants were asked to write group reports that were later integrated in a final report by an executive group composed of representatives of the individual groups. The final report was subsequently presented to the executive staff of the Census Bureau, as well as to the advisory committees of the American Statistical Association, the American Marketing Association, and the American Economic Association.

The Year 2000 Planning Program was successful on several counts. The report was received with moderate approval by all groups, and most members thought that the program should be continued in some form, perhaps permanently. Two creative products of the project were suggestions to establish an ombudsman to protect the interests of users of Census data and to create a Census University to develop and execute the bureau's continuing education programs. Those who were most positive about the report were oriented toward strategic concerns involving relatively ill-structured problems; less positive reactions came from persons with tactical or operational orientations toward well-structured problems. The program itself was made possible by a recognition among top-level bureau staff, including the director, that the bureau was confronted by important long-range problems whose structure was highly complex. Although the program involved significant resource allocations, there did not appear to be major risks in executing the program.

The main difference between brainstorming and other techniques for problem structuring is that the focus is on groups of knowledgeable individuals rather than those who

[73]Seyon H. Brown, "Scenarios in Systems Analysis," in *Systems Analysis and Policy Planning: Applications in Defense*, ed. E. S. Quade and W. I. Boucher (New York: American Elsevier Publishing, 1968), p. 305.
[74]Olaf Helmer, *Social Technology* (New York: Basic Books, 1966), p. 10. Quoted in Quade, *Analysis for Public Decisions*, p. 188.
[75]See Ian I. Mitroff, Vincent P. Barabba, and Ralph H. Kilmann, "The Application of Behavioral and Philosophical Technologies to Strategic Planning: A Case Study of a Large Federal Agency," *Management Science* 24, 1 (1977): 44–58.

have been identified, usually by means of ascription, as experts. Moreover, brainstorming activities are assessed not in terms of logical consistency or the plausibility of comparisons but according to consensus among members of a brainstorming group. The major limitation of consensus as a criterion of evaluation is that conflicts about the definition of problems may be suppressed, foreclosing opportunities to generate potentially appropriate ideas, goals, and strategies. While the Year 2000 Planning Program sought to create an open and permissive atmosphere, the final evaluation of the program's success was based on consensus among authoritative decision-makers (agency executive staff) and experts (advisory committees of professional associations). In short, this program provided no explicit procedures to promote the creative use of conflict in structuring problems.

Multiple Perspective Analysis

Multiple perspective analysis is a method for gaining insight into problems by applying three kinds of perspectives—personal, organizational, and technical—to problem situations.[76] Seen as an alternative to the near-exclusive emphasis on so-called rational-technical approaches, multiple perspective analysis is expressly designed to address ill-structured problems. Although there are many characteristics of each of the three perspectives, their major features are as follows:

1. *Technical perspective.* The technical (T) perspective views problems and solutions in terms of optimization models and employs microeconomics, benefit–cost analysis, systems analysis, and decision analysis. The technical perspective, said to be based on a scientific–technological worldview, emphasizes causal reasoning, objective analysis, prediction, and statistical inference. An example of the T perspective is the decision to drop the atomic bomb on Japan. The problem was seen to be composed of five alternatives: bombing and blockade, invasion, atomic attack without warning, atomic attack after warning, and dropping the bomb on an uninhabited island. Given the goal of unconditional surrender with a minimum loss of Allied lives and destruction of Japan, the T perspective resulted in the choice of the third alternative: atomic attack without warning.[77]
2. *Organizational perspective.* The organizational (O) perspective views problems and solutions as part of an orderly progression (with minor but temporary crises) from one organizational state to another. Standard operating procedures, rules, and institutional routines are major characteristics of the O perspective, which is often opposed to the T perspective and only minimally concerned with improving organizational performance. The decision to drop the atomic bomb shows how the O perspective differs from the T perspective. From an O perspective, a decision not to

[76]See Harold A. Linstone, *Multiple Perspectives for Decision Making: Bridging the Gap between Analysis and Action* (New York: North-Holland Publishing, 1984), and Linstone and others, "The Multiple Perspective Concept: With Applications to Technology Assessment and Other Decision Areas," *Technological Forecasting and Social Change* 20 (1981): 275–325.

[77]Results of analyses based on a technical (T) perspective may yield efficient answers. However, as a Japanese colleague writes, this does not mean that a moral (M) perspective may not yield more just, fair, or humane conclusions.

use the bomb raised political fears, because the $2 billion in funding for the atomic bomb was expended without congressional approval. Dropping the bomb showed Congress that the funds were not wasted. It also inaugurated the Cold War with the Soviet Union.

3. *Personal perspective.* The personal (P) perspective views problems and solutions in terms of individual perceptions, needs, and values. Characteristics of the personal perspective are an emphasis on intuition, leadership, and self-interest as factors governing problem structuring. The example of the atomic bomb shows how the P perspective supplies insights not available from either the T or the O perspectives. In 1945 the new President, Harry Truman, was an outsider to the FDR establishment, which had grown and solidified during Roosevelt's three terms in office. Truman lacked the legitimacy and influence necessary to challenge the establishment, including entrenched bureaucratic interests and policies, so early in his presidency. A decision not to drop the atomic bomb would be perceived as a sign of weakness to contemporaries and to future historians. Truman, who had a strong sense of history, wanted to appear as a bold and decisive leader.

Multiple perspective analysis is relevant to virtually any problem found in areas of public policy analysis, strategic planning, regional and urban analysis, and other domains. Linstone and colleagues have developed some of the following guidelines for employing multiple perspective analysis with groups:

- *Inter-paradigmatic mix.* Form groups on the basis of an inter-paradigmatic rather than interdisciplinary mix. For example, a team composed of a businessperson, lawyer, and writer is preferred to a team with an economist, a political scientist, and a psychologist. The inter-paradigmatic mix is preferable because it maximizes opportunities to find T, O, and P perspectives, whereas the more rigid boundaries of academic disciplines may constrain thinking.
- *Equal distribution of perspectives.* It is not possible to decide how much emphasis to place on T, O, and P perspectives at the outset of using multiple perspective analysis. As the group engages in its work, the discovery of the proper balance among the three perspectives will permit assignments to T, O, and P tasks. In the meantime, an equal distribution is preferable.
- *Uneven replicability.* The T perspective typically employs methods (e.g., experimental design) that are replicable. The O and P perspectives are not so easy to replicate. Much like a jury trial or a non-routine executive decision, the process is not as replicable as an experiment. The T, O, and P perspectives are replicable to different degrees.
- *Appropriate communications.* Adapt the medium of communication to the message. Summaries, oral briefings, scenarios, and vignettes are appropriate for communicating with those who hold O and P perspectives. Computer models, data sets, lists of variables, and analytical routines are appropriate for those with a T perspective.
- *Deferred integration.* Leave the integration of perspectives to the client or policymaker, but point out linkages among the T, O, and P perspectives and the varied conclusions they yield.

Multiple perspective analysis has been employed extensively in technology assessment and other areas of public policy. Methods of multiple perspective analysis, developed on the basis of earlier work in foreign policy analysis and the design of information systems,[78] is another problem-structuring method.

Assumptional Analysis

Assumptional analysis is a method that aims at the creative synthesis of conflicting assumptions about problem situations.[79] Assumptional analysis is relatively comprehensive, because it may be used with other methods and with groups, individuals, or both. Assumptional analysis is explicitly designed for problem situations where policy analysts, policymakers, and other stakeholders cannot agree on how to formulate a problem. The main criterion for assessing the adequacy of a given formulation of a problem is whether conflicting assumptions about a problem situation are surfaced, challenged, and synthesized.

Assumptional analysis is designed to overcome four major limitations of policy analysis: (1) policy analysis is often based on the assumption of a single decision-maker with clearly ordered but static preferences exercised at a single point in time; (2) policy analysis typically fails to consider in a systematic and explicit way strongly differing views about the nature of problems and their potential solutions; (3) most policy analysis is carried out in organizations whose "self-sealing" character makes it difficult or impossible to challenge prevailing formulations of problems; and (4) criteria used to assess the adequacy of problems and their solutions often deal with surface characteristics of problem formulations (e.g., logical consistency), rather than with basic assumptions underlying the conceptualization of problems.

Assumptional analysis explicitly recognizes the importance of conflict and commitment among members of a group. Maximally opposing policies are generated through conflict, a dialectical process that aids in uncovering and challenging the underlying assumptions that each policy makes. "Commitment on the other hand is also necessary if the proponents for each policy are to make the strongest possible case (not necessarily the best) for their respective points of view."[80] Assumptional analysis involves the use of five sequentially ordered methods (Figure 3.15).

1. *Stakeholder identification.* In the first phase, policy stakeholders are identified, ranked, and prioritized. The identification, ranking, and prioritization of stakeholders is based on an assessment of the degree to which they influence and are influenced by

[78]Multiple perspective analysis builds upon and extends previous works in foreign policy and systems design, specifically, Graham Allison, *Essence of Decision: Conceptual Models and the Cuban Missile Crisis* (Boston, MA: Little, Brown and Company, 1962), and C. West Churchman, *The Design of Inquiring Systems* (New York: Basic Books, 1971).

[79]See Ian I. Mitroff and James R. Emshoff, "On Strategic Assumption-Making: A Dialectical Approach to Policy and Planning," *Academy of Management Review* 4, 1 (1979): 1–12; Richard O. Mason and Ian I. Mitroff, *Challenging Strategic Planning Assumptions: Theory, Cases, and Techniques* (New York: Wiley, 1981); and Ian I. Mitroff, Richard O. Mason, and Vincent P. Barabba, *The 1980 Census: Policymaking amid Turbulence* (Lexington, MA: D. C. Heath, 1983).

[80]Mitroff and Emshoff, "On Strategic Assumption-Making," p. 5.

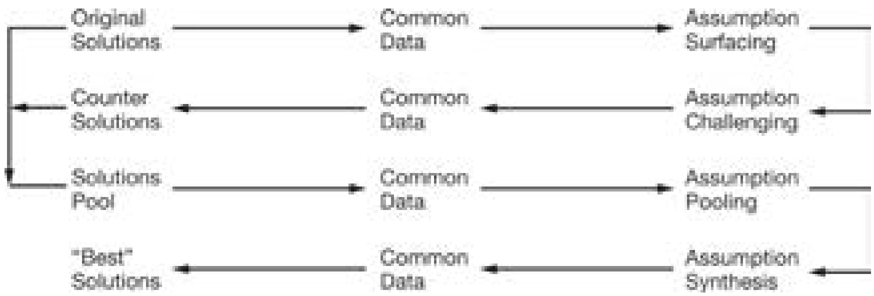


FIGURE 3.15
The Process of Assumptional Analysis

Source: Adapted from Mitroff and Emshoff (1979).

the policy process. This procedure results in the identification of stakeholders—for example, dissident groups of administrators or clients—who are usually excluded.

2. *Assumption surfacing.* In the second phase, analysts work backward from a recommended solution for a problem to the selective sources of data that support the recommendation, as well as the underlying assumptions that, when coupled with the data, allow one to deduce the recommendation. Each recommended solution should contain a list of assumptions that explicitly and implicitly underlie the recommendation. By listing all assumptions—for example, that poverty is a consequence of historical accidents, elite domination, unemployment, cultural deprivation, and so on—there is an explicit specification of the problem to which each recommendation is addressed.
3. *Assumption challenging.* In the third phase, analysts compare, contrast, and evaluate recommendations and their underlying assumptions. This is done by systematically comparing assumptions and counter-assumptions that differ maximally. During this process, each assumption is challenged by a counter-assumption. If a counter-assumption is implausible, it is eliminated from further consideration. If it is plausible, it is examined to determine whether it might serve as a basis for a new conceptualization of the problem and its solution.
4. *Assumption pooling.* When the assumption-challenging phase has been completed, the solutions generated in previous phases are pooled. Here assumptions (rather than recommendations) are prioritized according to their relative certainty and importance to different stakeholders. Only the most important and uncertain assumptions are pooled. The ultimate aim is to create an acceptable list of assumptions on which as many stakeholders as possible agree.
5. *Assumption synthesis.* The final phase is the creation of a composite or synthetic solution for the problem. The composite set of acceptable assumptions can serve as a basis for the creation of a new conceptualization of the problem. When issues surrounding the conceptualization of the problem and its potential solutions have reached this point, the activities of stakeholders may become cooperative and cumulatively productive.

The four phases of assumptional analysis are illustrated in Figure 3.15, which helps visualize important features of the method. First, the method begins with recommended solutions for problems rather than assumptions themselves. This is because most policy stakeholders are aware of proposed solutions for problems but seldom conscious of underlying assumptions. By starting with recommended solutions, the method builds on what is most familiar to stakeholders but then goes on to use familiar solutions as a point of reference for forcing an explicit consideration of underlying assumptions. A second important feature of the technique is that it attempts, as far as possible, to focus on the *same* set of data or policy-relevant information. The reason for this is that conflicts surrounding the conceptualization of policy problems are not so much matters of "fact" but matters involving conflicting interpretations of the same data. Although data, assumptions, and recommended solutions are interrelated, it is not so much the problem situation (data) that governs the conceptualization of problems but the assumptions that analysts and other stakeholders bring to the problem situation. Finally, assumptional analysis systematically addresses a major problem of policy analysis, which is that of applying systematic procedures to deal with conflict.

The aims and procedures of assumptional analysis are closely related to the modes of policy argument presented in chapters 1 and 8.[81] Each mode of policy argument is based on different kinds of assumptions that are used to produce alternative conceptualizations of problems. Assumptional analysis is therefore a major vehicle for conducting reasoned debates about the nature of policy problems. Assumptional analysis may be used with groups of policy stakeholders who actually participate in structuring policy problems or by an individual analyst who simulates the assumptions of stakeholders. Assumptional analysis can help reduce errors of the third type (E_{III}).

Argument Mapping

In *argument mapping,* the plausibility and importance of elements of a policy argument are presented as graphic displays. The first step is to rate different elements—information, warrants, backings, objections, and rebuttals—on plausibility and importance scales. Take the warrant of the argument to abandon the 55 mph speed limit (National Maximum Speed Law of 1974). This particular argument has been based on the warrant that the opportunity costs of time that is lost driving at slower speeds results in increased speeds. In turn, increased speeds and riskier driving occurs among motorists with higher incomes. Conversely, it might be argued that there are fewer accidents among younger motorists, a claim that is based on the warrant that younger drivers earn less income, have lower opportunity costs, take fewer risks, and therefore have fewer accidents. This warrant may be rated by different stakeholders on nine-point plausibility and importance scales (1 = low, 9 = high) and plotted on a graph such as that shown in Figure 3.16.[82]

---

[81]For an elaboration of assumptional analysis using the structural model of argument presented in Chapters 1 and 8, see Mitroff, Mason, and Barabba, *The 1980 Census, passim.*

[82]Computer software called *Rationale 2* provides capabilities for entering, editing, saving, and evaluating complex policy arguments. The program provides color codes, shading, and a 3-point scale to rate each element of an argument. See www.austhink.com.

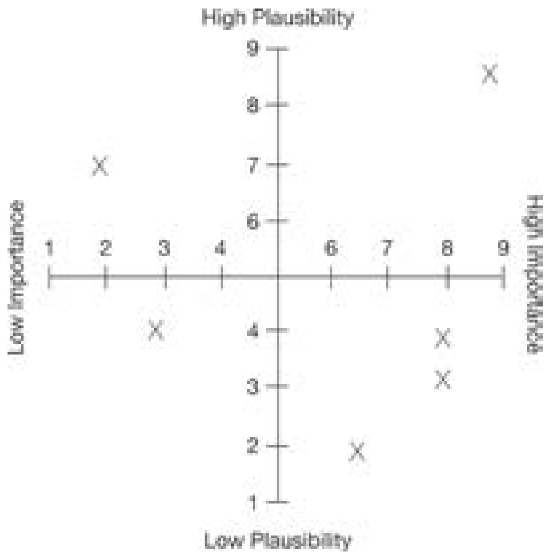


FIGURE 3.16
Plausibility and Importance of Warrant about Opportunity Costs of Driving at Lower Speeds

Figure 3.16 displays the plausibility and importance ratings for six stakeholders. The graph shows that stakeholders are distributed across all four quadrants, indicating broad disagreement about the plausibility and importance of the warrant, which would no doubt generate objections and rebuttals. When stakeholders are members of a particular group, the disagreements plotted on the right half of the graph (i.e., high importance) can be discussed and perhaps resolved. The typical situation, however, is one where the analyst must identify a range of stakeholders on the basis of telephone interviews, documents, and websites containing the stakeholders' arguments and assumptions. Judgments are then made about the plausibility and importance that stakeholders attach to a warrant, backing, objection, or rebuttal. For example, a review of documents on the 55 mph speed limit indicates that the stakeholder rating the warrant as highly plausible ($P = 9$) is an economist, whereas the stakeholder who ranks the warrant as having low plausibility ($P = 2$) is an ethnographer who specializes in the culture of young drivers, basing his analyses on ethnographic interviews with young drivers, parents, and law enforcement personnel.[83]

On the basis of arguments and evidence provided by these ethnographic and economic sources, supplemented with information about the reasons provided by stakeholders and statistics on accident rates by age group—we would conclude that this particular warrant has low plausibility. However, because the warrant has high importance it is relevant to the conclusions of the argument and must be retained for further examination.

[83]See Thomas H. Forrester, Robert F. McNown, and Larry D. Singell, "A Cost-Benefit Analysis of the 55 mph Speed Limit," *Southern Economic Journal* 50 (1984), reviewed by George M. Guess and Paul G. Farnham, *Cases in Public Policy Analysis* (New York: Longman, 1989), p. 199. For an ethnographic analysis of younger drivers and the meanings they attach to accidents and fatalities—which are different from those predicted by economic theory—see J. Peter Rothe, *Challenging the Old Order* (New Brunswick, NJ: Transaction Books, 1990).

CHAPTER SUMMARY

This chapter has provided an overview of the nature of policy problems, described the process of structuring these problems, examined relationships among policy models, and described different methods of problem structuring. The chapter shows that one of the most important challenges facing policy analysts is to reduce the likelihood of a Type III error: defining the wrong problem.

REVIEW QUESTIONS

1. "Our problem is not to do what is right," stated Lyndon Johnson during his years in the White House. "Our problem is to know what is right." Considering the characteristics of policy problems discussed in this chapter, to what extent can we know which policy is the "right" one?
2. A commonly accepted viewpoint among analysts in government and in universities is that policy analysis can be objective, neutral, and impartial. Considering the characteristics of ill-structured problems, to what extent is this viewpoint plausible?
3. Provide two or three examples from your own experience of ways that worldviews, ideologies, or popular myths shape the formulation of policy problems.
4. There are several broad types of organizational structures in which policy formation occurs. One type is the "bureaucratic" structure, whose characteristics include centralization, hierarchical chain of command, specialization of tasks, and complete information. The bureaucratic form of organization requires consensus on preferred policy outcomes, as well as certainty that alternative courses of action will result in certain preferred outcomes (J. D. Thompson, *Organizations in Action* (New York: McGraw-Hill, 1967), pp. 134–135). If many of our most important policy problems are ill-structured or "wicked," how appropriate is the bureaucratic form of organization for formulating and resolving such problems?
5. If many of our most important problems are relatively ill-structured or wicked, to what extent is it possible to hold policymakers politically and morally responsible for their actions? See M. M. Webber and H.W.J. Rittel, "Dilemmas in a General Theory of Planning," *Policy Sciences* 4, 2 (1973): 155–169.
6. Construct a scenario on the state of one of the following problem situations in the year 2050:
   Availability of public mass transit
   Arms control and national security
   Terrorism and public safety
   Quality of the public school system
   State of the world's ecological system
7. Select two editorials on a current issue of public policy from newspapers (e.g., *The New York Times, The Washington Post, The Economist, Le Monde*), news magazine (e.g., *Newsweek, The New Republic, National Review, CATO Institute*), or websites (e.g., www.aarp.org, www.cap.org). After reading the editorial:
   a. Use the procedures for argumentation analysis (Chapter 1) to display information, warrants, backings, objections, and rebuttals for two *competing arguments* about the issue contained in the editorials.
   b. Rate the warrants for each argument and plot them according to their plausibility and importance (Figure 3.16).
   c. Which argument is the most persuasive? Why?

DEMONSTRATION EXERCISES

1. Choose a policy issue area such as crime control, national security, environmental protection, economic development, or choose your own area. Use the procedures for stakeholder

analysis presented in Exhibit 3.1 to generate a list of stakeholders who affect or are affected by problems in the issue area you have chosen. After generating the list, create a cumulative frequency distribution. Place stakeholders on the horizontal axis, numbering them 1 . . . *n*. On the vertical axis, place the number of new (nonduplicate) ideas generated by each stakeholder (the ideas might refer to objectives, alternatives, outcomes, causes, etc.). Connect the number of new ideas of each stakeholder with a line graph.
- Does the line graph flatten out?
- If so, after how many stakeholders?
- What conclusions do you draw about the policy problem(s) in the issue area?

Compare your work with Case 3.1 at the end of the chapter.

2. After reading Case 3.1, write an essay in which you compare and contrast the process of boundary analysis and estimation in the three cases. In your comparison, address these questions:
   - What are the key differences in data collection, represented by the process of group interviewing and content analysis?
   - Why do the cumulative frequency graphs flatten out the way they do?
   - Evaluate the statement: "Boundary analysis is a reliable way to estimate the 'universe of meanings' in a given policy issue area."

BIBLIOGRAPHY

Ackoff, Russell L. *The Art of Problem Solving: Accompanied by Ackoff's Fables*. New York: John Wiley, 1978.

Ackoff, Russell L. *Redesigning the Future*. New York: Wiley, 1974.

Adams, James L. *Conceptual Blockbusting*. Stanford, CA: Stanford Alumni Association, 1974.

Batie, Sandra S. "Wicked Problems and Applied Economics." *American Journal of Agricultural Economics* 90, 5 (2008): 1176–1191.

Bobrow, Davis B., and John S. Dryzek. *Policy Analysis by Design*. Pittsburgh, PA: University of Pittsburgh Press, 1986.

Campbell, Donald T. "Pattern Matching as an Essential in Distal Knowing," pp. 81–106 in Kenneth R. Hammond (ed.), *The Psychology of Egon Brunswik*. New York: Holt, Rinehart, and Winston, 1966.

Churchman, C. West. *The Design of Inquiring Systems*. New York: Basic Books, 1971.

Dery, David. *Problem Definition in Policy Analysis*. Lawrence: University Press of Kansas, 1984.

Dror, Yehezkel. *Public Policy Making Reexamined*. Rev. ed. New Brunswick. NJ: Transaction Books, 1983.

Dryzek, John S. "Policy Analysis as a Hermeneutic Activity." *Policy Sciences* 14 (1982): 309–329.

Dunn, William N. "Methods of the Second Type: Coping with the Wilderness of Conventional Policy Analysis." *Policy Studies Review* 7, 4 (1988): 720–737.

Dunn, William N. "A Pragmatic Strategy for Discovering and Testing Threats to the Validity of Sociotechnical Experiments." *Simulation Modelling Practice and Theory* 10, 3 (2002): 169–194.

Dunn, William N. "Pragmatic Eliminative Induction: Proximal Range and Context Validation in Applied Social Experimentation." *Philosophica-Gent* (1997): 75–112.

Dunn, William N. "Using the Method of Context Validation to Mitigate Type III Errors in Environmental Policy Analysis," pp. 417–436 in *Knowledge, Power and Participation In Environmental Policy Analysis*. New Brunswick, NJ: Transaction, 2001.

Fischhoff, Baruch. "Clinical Policy Analysis," in William N. Dunn (ed.), *Policy Analysis: Perspectives, Concepts, and Methods*. Greenwich, CT: JAI Press, 1986.

Fischhoff, Baruch. "Cost-Benefit Analysis and the Art of Motorcycle Maintenance." *Policy Sciences* 8 (1977): 177–202.

George, Alexander. "Criteria for Evaluation of Foreign Policy Decision Making." *Global Perspectives* 2 (1984): 58–69.

Hammond, Kenneth R. "Introduction to Brunswikian Theory and Methods." *New Directions for Methodology of Social and Behavioral Science* 3 (1980): 1–12.

Hammond, Kenneth R. "Probabilistic Functionalism: Egon Brunswik's Integration of the History, Theory, and Method of Psychology," pp. 15–81 in K. R. Hammond (ed.), *The Psychology of Egon Brunswik* (New York: Holt, Rinehart, and Winston, 1966).

Hofstadter, Richard. *Gödel, Escher, Bach*. New York: Random House, 1979.
Hogwood, Brian W., and B. Guy Peters. *The Pathology of Public Policy*. Oxford: Clarendon Press, 1985.
Linder, Stephen H., and B. Guy Peters. "A Metatheoretic Analysis of Policy Design," in William N. Dunn and Rita Mae Kelly (eds.), *Advances in Policy Studies since 1950*. Vol. 10 of *Policy Studies Review Annual*. New Brunswick, NJ: Transaction Books, 1992.
Linstone, Harold A. *Multiple Perspectives for Decision Making*. New York: North-Holland, 1984.
Mason, Richard O., and Ian I. Mitroff. *Challenging Strategic Planning Assumptions*. New York: Wiley, 1981.
Mitroff, Ian I., and James R. Emshoff. "On Strategic Assumption-Making: A Dialectical Approach to Policy and Planning." *Academy of Management Review* 4, 1 (1979): 1–12.
Mitroff, Ian I., Richard O. Mason, and Vincent P. Barabba. *The 1980 Census: Policymaking Amid Turbulence*. Lexington, MA: D. C. Heath, 1983.
O'Shaughnessy, John. *Inquiry and Decision*. New York: Harper & Row, 1973.
Rittel, Horst W. J. and Melvin M. Webber, "Dilemmas in a General Theory of Planning," *Policy Sciences* 4 (1973): 155–169.
Saaty, Thomas L. *The Analytic Hierarchy Process*. New York: McGraw-Hill, 1980.
Schon, Donald A. *The Reflective Practitioner*. New York: Basic Books, 1983.
Sieber, Sam. *Fatal Remedies*. New York: Plenum Press, 1981.
Warfield, John N. *Societal Systems: Planning, Policy, and Complexity*. New York: Wiley, 1976.
Watzlawick, Paul, John Weakland, and Richard Fisch. *Change: Principles of Problem Formation and Problem Resolution*. New York: W. W. Norton, 1974.

EXHIBIT 3.1 POLICY STAKEHOLDER ANALYSIS: THE AFFORDABLE CARE ACT (OBAMACARE)

Overview

A stakeholder is a person who speaks for or represents a group that is affected by or affects a problem. Stakeholders include the Speaker of the U.S. House of Representatives, the head of a legislative committee, an executive director of an advocacy group, or membership and advocacy organizations such as the American Association for Retired Persons, the National Rifle Association, Human Rights Watch, the National Welfare Rights Organization, or the Grey Panthers. Policy analysts are also stakeholders, as are the organizations that hire them as consultants or appoint them to positions in NGOs or government. A person who speaks for or represents a group or organization that is not affected by, or does not affect a problem, does not have a stake in this particular problem. Therefore s/he is not a policy stakeholder. Many news reporters or uninvolved college professors are *not* stakeholders.

Guidelines for Identifying the Policies of Stakeholders

- Stakeholders are best identified by policy issue area. A policy issue area is a set of policy issues in a subject area, for example, there are policy issues in education, housing, welfare, employment, personal rights, and international security. There are more than forty policy areas in many countries. In education, for example, stakeholders in the United States include state and federal departments of education, teachers' unions, parent-teacher organizations, library associations, and student groups. In the issue area of health policy there are literally thousands of organizations conducting research, representing the professional interests of their members, or advocating policy positions favored by their members.
- Stakeholders should be identified by specific names and titles. For example, in the issue area of health, a number of organizations have a

stake in the Patient Protection and Affordable Care Act of 2010 (Obamacare). These include the American Medical Association, the American Cancer Society, the American Association of Retired Employees, the American Hospital Association, the Center for Medicaid and Medicare Services (CMMS) of the Department of Health and Human Services, Pharma, and a variety of organizations that sell equipment and software to state and federal governments, hospitals, and physicians (Figure 3.C.1). Typically, each of these stakeholders has a distinctive position on health policies.

- A sociometric (also called "snowball") sampling procedure of stakeholders such as that described below is an effective way to estimate the population of stakeholders who affect and are affected by policies in a particular issue area.

Step 1: Use Google or a reference book such as *The Encyclopedia of Associations (EA),* which has some 35,000 research, membership, and advocacy organizations at national, subnational, and international levels in the United States and abroad. Use the search utility in *EA* to identify and list twenty top stakeholder organizations (by membership or operating budget) that have taken a position on a policy in one of the seventeen issues areas covered by the *Encyclopedia*. These issue areas include trade, agriculture, environment, science, technology and engineering, education, social welfare, and health.

Step 2: For each stakeholder, obtain a policy document (e.g., a report or record of congressional testimony) and list specific policies the organization supports and opposes.

Step 3: List the policies of the first stakeholder, the second stakeholder, the third, and so on. Do *not* count duplicates.

Step 4: Draw a graph that displays on the vertical axis unique (non-duplicate) policies for each of the twenty stakeholders. The graph should gradually flatten out, with no new policies mentioned. If the graph does not flatten out before reaching the twentieth stakeholder, repeat steps 1–4, adding new stakeholders and policies. If it flattens out before the twentieth stakeholder, stop the collection of policies.

Step 5: (Optional) Add to the sample ten governmental stakeholders who should be listed because of their formal positions. Repeat Steps 2 through 4, listing only the *new* policies of the governmental stakeholders. You now have an estimate of the "population" of policies of the key stakeholders who affect and are affected by Obamacare, along with a description of the positions of stakeholders on the policies. This is a helpful for structuring the problem. ■

CASE 3.1 STRUCTURING PROBLEMS OF RISK: SAFETY IN MINES AND ON HIGHWAYS

Relatively ill-structured (or so-called "wicked") problems must be structured before they can be solved. The process of structuring a problem requires the search for, identification, and arrangement of the "elements" that constitute a problem. A problem element refers to one or more of the following:

Stakeholders. Which stakeholders affect or are affected by a problem situation,[84] as perceived by the stakeholders?

Alternatives. What alternative courses of action may be taken to solve the perceived problem?

Actions. What actions should be taken to adopt the alternatives?

Outcomes. What are the outcomes of actions that may be taken to initiate the alternatives?

Values (Utilities). What is the value of the outcomes?

Many of the most practically significant problems in the world today are ill-structured or wicked. One or more of the above elements may be improperly omitted in the process of problem structuring (this is omitted variable bias). Even when virtually all or most problem elements have been correctly identified, relations among the elements may be complex or ambiguous. Among other things, this makes it difficult or impossible to determine the strength and significance, practical as well as statistical, of causal relations among elements. For example, global warming is usually seen as an ill-structured or wicked problem, as are hunger and poverty. In the case of global warming, many causal processes that are believed to link atmospheric pollution, global warming, and climate change are complex and obscure. An essential point is that the complexity of these processes stems not only from the biological and other processes of nature, but also from the different and often conflicting beliefs of stakeholders who disagree about the formulation of the problem and its potential solutions. In this context, the possible combinations and permutations of stakeholders, alternatives, actions, outcomes, and values usually looks *unmanageably huge*.

Under these conditions, and contrary to Simon's belief that new and unprecedented methods are *not* needed to structure ill-structured or wicked problems, it appears that many conventional methods of decision analysis, microeconomics, and political analysis are of limited value until an ill-structured problem has been "tamed" by means of problem-structuring methods whose role is to convert ill-structured into well-structured problems. To be sure, conventional methods are useful in

[84]A *problem situation,* following pragmatists John Dewey and Charles Sanders Peirce, becomes a *problem* only after it has been structured (another more general term is formulated). Problem structures range from relatively well structured and moderately structured to ill structured. An ill-structured problem may become a moderately structured or well-structured problem in the course of a problem-structuring process involving the search for information. See Herbert A. Simon, "The Structure of Ill Structured Problems," *Artificial intelligence* 4, 3–4 (1973): 181–201. Simon (p. 183) notes that "there is no real boundary between WSPs and ISPs, and no reason to think that new and hitherto unknown types of problem solving processes are needed to enable artificial [or natural] intelligence systems to solve problems that are ill structured." Recall that a problem situation is a diffuse set of worries and inchoate signs of stress, which is similar to unprocessed data that must be interpreted before concepts may be formed and applied. Simon is wrong, however, in claiming that no special problem-solving processes are required simply because the two kinds of problems are not mutually exclusive and disjoint. The range of pollutants in agricultural societies and industrial societies calls for different analytical methods and practical interventions, and it is often useful to create artificial dichotomies to partition relatively "wicked" from relatively "tame problems."

solving relatively well-structured problems (e.g., deterministic ones, where probability plays no role) about which we are certain, for example, problems involving fixed quantities in a spreadsheet that we must sum, multiply, or divide. Conventional methods are also useful in solving moderately structured (probabilistic) problems involving the estimation of uncertainty, for example, estimating the uncertainty of outcomes within a confidence interval.

However, ill-structured problems occupy a different order.[85] Estimates of uncertainty cannot be made with ill-structured problems because we do not know the outcomes to which we might attach probabilities. Here, the analyst is much like an architect who has been commissioned to design a *custom* house for which there is no *standard* plan.[86] The adoption of a conventional plan of the kind used in constructing a tract home would be a failure.

Public policies are deliberate attempts to change social systems. The process of formulating and implementing policies occurs in social systems where many contingencies lie beyond the control of policymakers. It is these unmanageable contingencies that are usually responsible for the success and failure of policies in achieving their objectives. The "contingencies" (unforeseen events) may be seen as potential rival hypotheses that may challenge the claim that a policy "worked." In such cases, it is desirable to test, and where possible eliminate, these rival hypotheses (contingencies) through a process of *eliminative* induction. Eliminative induction takes the general form: "Repeated observations of policy *x* and outcome *y* confirm that *x* is causally relevant to the occurrence of *y*. However, additional observations of *x, z,* and *y* confirm that unmanageable contingency, *z,* and not policy, *x,* is responsible, or primarily responsible, for the occurrence of *y*." By contrast, *enumerative* induction takes the general form: "Repeated observations confirm that the policy *x* is causally relevant to the occurrence of policy outcome *y*." As humans, we naturally seek out events that confirm what we believe. We do not like surprises, although they occur frequently.

Eliminative induction permits us to identify and explain surprises that do not conform to our expectations. Because the number of contingencies is usually very large, the process of identifying and testing rival explanations is never complete. It seems impossible to identify and test an unmanageably huge number of potential rival hypotheses. How can this be done?

One answer is creativity and imagination. But creativity and imagination are very difficult to teach or learn, because the rules governing creative or imaginative solutions are vague. Another answer is an appeal to well-established theories. However, theories in the social sciences are disputed and controversial.

One answer is the use of *boundary analysis* to identify contingencies and arrange them, that is, give them *structure*. This structured arrangement of elements may be seen to involve a large number of rival hypotheses. Boundary analysis looks for rival hypotheses in the naturally occurring policy debates that take place among stakeholders. The aim of boundary analysis is to obtain a relatively complete set of rival hypotheses in a given policy context. Although boundary estimation strives to be comprehensive, it does not attempt the hopeless task of identifying and testing *all* plausible rival hypotheses, any more than a sample drawn from a population strives to obtain *all* elements sampled—the Central Limit Theorem tells us when we have a sufficient number of elements, depending on the level of statistical significance and certainty we wish to achieve. Thus, although the full range of rival hypotheses is never complete, it is possible to *estimate* the boundaries of the problem.

(*continued*)

[85]Elsewhere I have called attention to the intersection of types of methods and types of problems, arguing that ill-structured problems require for their solution nonconventional "methods of the second type," for example, heuristics such as brainstorming that permit an open-ended search for problem elements, or what Simon generalizes as information. Dunn, "Methods of the Second Type."

[86]The architecture analogy is from Herbert Simon in "The Structure of Ill Structured Problems."

Assessing the Impact of National Maximum Speed Limit

The author conducted a boundary analysis by using documents prepared by thirty-eight state officials responsible for reporting on the effects of the original 55 mph speed limit, passed into law by Congress in 1974. The documents also reported the effects of the 1984–1985 speed limits that raised the maximum speed limit in forty states to an average of approximately 65mph.

There were sharp disagreements among many of the thirty-eight stakeholders. For example, stakeholders in some states (Pennsylvania and New Jersey) were firmly committed to the belief that speed limits were causally related to a decline in highway fatalities and injuries. Other stakeholders (e.g., Illinois, Washington, Idaho) were just as firmly committed to the belief that speed limits had no effect, or a negligible effect, on fatalities and injuries.

Of great importance to the problem-structuring method of boundary estimation used in this case is that 718 plausible rival hypotheses about the causal relation between speeding and fatalities were used by thirty-eight stakeholders to affirm or dispute the effectiveness of the 55 mph speed limit in saving lives. Of this total, 109 causal hypotheses were unique, in that they did not duplicate hypotheses advanced by any other stakeholder.

When interpreting Figure 3.C.1 it is important to understand that, from the standpoint of communications theory and language, the information-content of a hypothesis is inversely proportional to the probability of the information being mentioned. Hypotheses that are mentioned more frequently—those on which there is greater consensus—have *less probative value* than rarely mentioned hypotheses. Highly probable or predictable hypotheses provide little or no new information.

The rival hypotheses in Figure C.3.1 represent the cumulative frequency of *unique* (nonduplicate) causal hypotheses. As the figure shows, the cumulative frequency curve of unique hypotheses begins to flatten out beginning with the twenty-second stakeholder. Although the *total* number

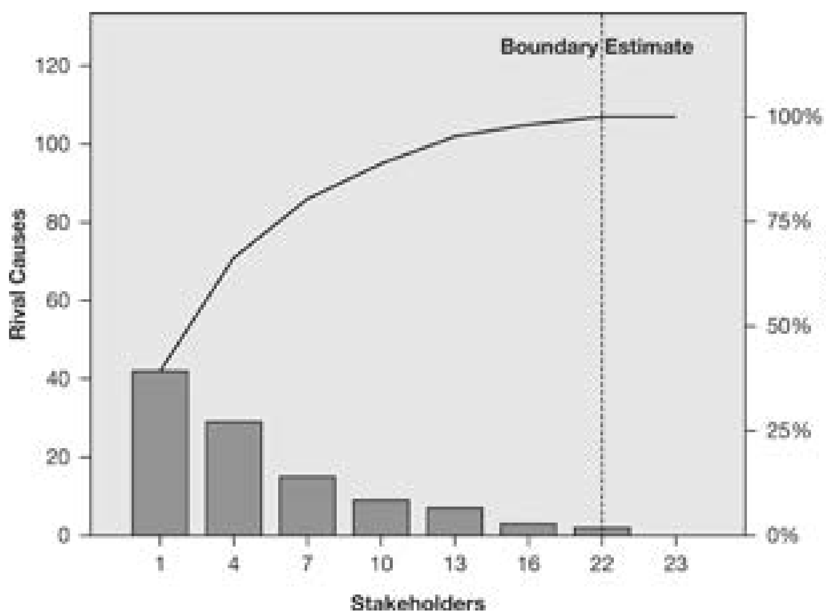


FIGURE C3.1
Pareto Chart—Cumulative Frequency of Rival Causes

of rival hypotheses continues to increase without *apparent* limit, the boundary of *unique* rival hypotheses is reached within a small number of observations. This indicates that a basically satisfactory definition of the elements of the problem has probably been achieved. It is important to note that, of the 109 unique rival hypotheses, several are related to the state of the economy—rate of unemployment, value of the international price of oil, level of industrial production—and these were used to explain the rise and fall of traffic fatalities, rather than the conventional wisdom that highway speeds and the mitigating effects of the 55 mph speed limit are decisive. A regression analysis based on economic factors plus the speed limit suggested that only 15 percent of fatalities were due to the speed limit.

Evaluating Research on Risk in Mine Safety and Health

In 1997, a branch of the U.S. Office of Mine Safety and Health Research began a process of strategic planning. The aim of the process was to reach consensus, if possible, on the prioritization of research that addresses different aspects of risk associated with the health and safety of miners.

Priority setting in this and other research organizations typically seeks to build consensus under conditions where researchers, research managers, and external stakeholders use conflicting criteria to evaluate the relative merits of their own and other research projects. Even when reliable and valid data are available—for example, quantitative data from large-sample studies of the probabilities of different kinds of mine injuries and deaths—it is often unclear whether "objective" measures of risk, by themselves, provide a sufficient basis for prioritizing research.

The difficulty is that extra-scientific as well as scientific factors affect judgments about the relative merits of research on risk. For example, "black lung" disease and other chronic conditions have been intensively investigated. Despite the importance of the black lung problem, additional research is not a priority. Accordingly, data on high expected severity (probability × severity) do not alone provide a sufficient basis for prioritizing research problems. These same high-risk accidents, however, might be a research priority if the aim is to translate findings into research on technologies for improving mine safety and health. Because judgments about research priorities are based on multiple, hidden, and frequently conflicting criteria, it is important to uncover and evaluate these criteria as part of the process of problem structuring. This is a classic ill-structured problem where the fatal error is defining the wrong problem.

The problem-structuring process had three major objectives. The first of these was to uncover hidden sources of agreement and disagreement, recognizing that *disagreement* is an opportunity for identifying alternative approaches to priority setting. Second, the process was designed to generate from stakeholders the *criteria* they use to evaluate research on risks affecting mine safety and health. Third, the process employed graphs, matrices, and other visual displays in order to *externalize* the criteria underlying individual judgments. The priority-setting process enabled each team member to understand and debate the varied reasons underlying the priorities used to evaluate research on risk.

Stakeholders were prompted to state the criteria they use to distinguish among twenty-five accident and health problems. Each stakeholder was presented with a form listing pre-randomized sets of three problems. One stakeholder was chosen at random to present criteria to the group. The first presented fourteen different criteria for evaluating research on risk, which included two criteria—severity and catastrophic potential of accidents and diseases—which were based on large-sample relative frequency (probability) data. The second randomly chosen team member presented seventeen additional (new) criteria; the third team member, eleven criteria; the fourth, twelve criteria, and so on, until no additional criteria could be offered that were not mentioned before. In all, eighty-four criteria were generated, none after the eleventh presenter. The approximate boundary of the problem was displayed with a Pareto chart to show the cumulative frequency of constructs (Figure C3.2).

The process of interactive group priority setting not only addressed the "objective" side of risk research but also captured "subjective" dimensions that go by such labels as "perceived risk," "acceptable risk," and "actionable risk."

(*continued*)

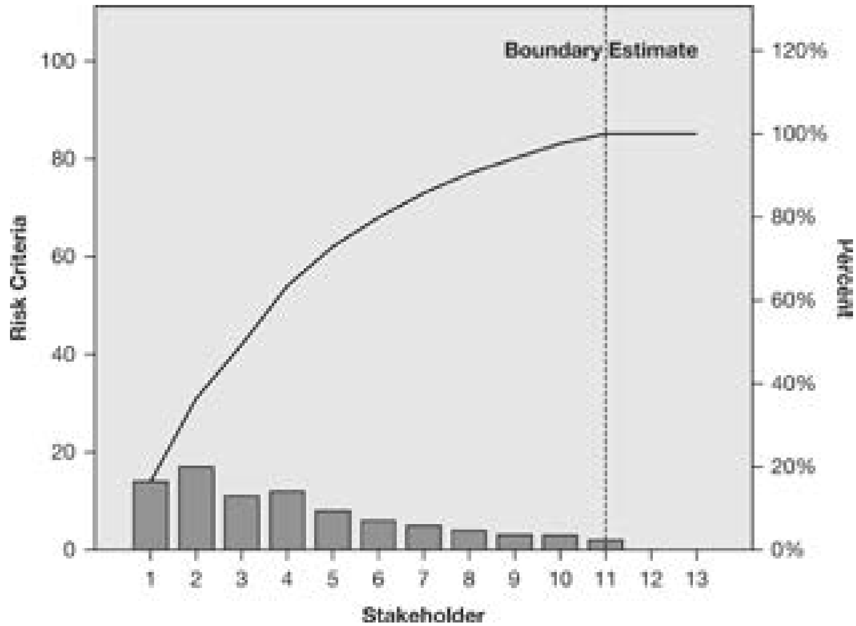


FIGURE C3.2
Pareto Chart—Cumulative Frequency of Criteria for Evaluating Research on Risk

The boundaries of the problem were estimated through an open process of generating and discussing dimensions of the problem that were otherwise tacit or concealed. The method of boundary analysis and estimation provides a "stop rule" which precludes an infinite search for dimensions of problems. ■

CASE 3.2 BRAINSTORMING ACROSS DISCIPLINES: THE ELEVATOR PROBLEM

The following story is based on a story told by Russell Ackoff, a master at solving ill-structured problems, or what he called "messes."[87] The story, a classic of unknown origin which has been in circulation for many years, shows why methods such as benefit–cost analysis, while appropriate for well-structured and moderately structured problems, are not appropriate for ill-structured problems.

[87]Russell Ackoff, *The Art of Problem Solving* (New York: Wiley-Interscience, 1978), "Fable 3.2 An Ups-and-Downs Story," pp. 53–54.

The manager of a large office building had been receiving an increasing number of complaints about the elevator service, particularly during rush hours. When office tenants threatened to move out, the manager decided that a solution must be found.

The manager called on a group of consulting engineers who specialized in the design and construction of elevators. After diagnosing the problem, the engineers identified three options: (1) add more elevators, (2) replace the existing elevators with faster ones, or (3) add a computerized control system so that elevator service could be optimally routed for faster service.

In collaboration with an economist, the engineers conducted a benefit–cost analysis of the three options. They found that investing in additional elevators, replacing the current elevators with faster models, or installing a computer control system would yield significant improvements of service. However, they found that the cost of any of the three alternatives was not justified by the income generated by renting the building. None of the alternatives was acceptable.

The manager called an urgent meeting of his staff and presented the problem to them as part of a *brainstorming* session similar to the one described earlier in this chapter. Many suggestions were made and then discarded. The discussion slowed down and stalled.

During a break a young assistant in the human resources department, who had been quiet to this point, made a suggestion which was eventually accepted by everyone. A few weeks later the manager made a modest investment, and the problem was solved.

Full-length mirrors were installed on all the walls of the elevator lobbies on each floor.

The young personnel psychologist observed that the waiting times were actually quite short. He correctly identified the problem—the complaints came from the boredom of waiting for the elevators. Yet the time only seemed long. He gave people the opportunity to look at themselves and others in the mirror (particularly those of the opposite sex) without appearing to do so.

The consulting engineers and the economist had committed a Type III error: they defined the wrong problem. Because of this they falsely concluded that there was no solution.[88] ■

[88]Another example of a Type III error is the aphorism popularized by Aaron Wildavsky: "No solution, no problem." See Wildavsky's "Preface" to *Speaking Truth to Power: The Art and Craft of Policy Analysis* (Boston, MA: Little Brown, 1979).

CHAPTER

4

Forecasting Expected Policy Outcomes

LEARNING OBJECTIVES

By studying the chapter, you should be able to:

- Distinguish extrapolation, prediction, and expert judgment
- Understand temporal, historical, and institutional effects on forecast accuracy
- Contrast potential, plausible, and normative futures
- Assess the accuracy of extrapolative, predictive, and judgmental forecasting methods
- Use statistical software to make point and interval estimates of expected policy outcomes
- Analyze cases that illustrate conceptual and technical issues in policy forecasting

INTRODUCTION

Forecasting provides a prospective vision of policy outcomes, enlarging capacities for understanding, control, and societal guidance. Forecasts, whether based on expert judgment, on the extrapolation of historical trend, or on technically sophisticated econometric models, are prone to errors based on faulty or implausible assumptions; the unrecognized effects of institutional incentive systems; and on the complexity of policy issues in areas ranging from health, welfare, and education to science, technology, and the environment.

We begin this chapter with an overview of the forms, functions, and performance of forecasting in policy analysis, stressing a range of criteria for assessing the strengths and limitations of different forecasting methods. We then compare and contrast extrapolative, theoretical, and judgmental forecasting methods and their role in producing information about expected policy outcomes. We conclude with a presentation of methods that combine the three approaches.

FORECASTING IN POLICY ANALYSIS

Forecasting is a set of procedures for creating information about future states of society on the basis of present and prior information. Forecasts take three principal forms: extrapolations, predictions, and expert judgments.[1]

1. An *extrapolation* is a forecast based on the projection of current and historical trends into the future. Projections often raise questions about the validity of conclusions based on information produced by the extrapolation of past trends into the future, for example, through time-series analysis. Projections are sometimes supplemented by arguments from authority, for example, the opinions of experts.
2. A *prediction* is a forecast based on theoretical explanations of why past trends should repeat themselves in the future. For example, explanations based on utility theory in economics that predict that money will become more valuable in future, but at a declining rate. As this example suggests, theoretical laws such as the law of diminishing utility of money may explain as well as predict the generative mechanisms ("causes") and consequences ("effects") required of predictions. Predictions may also be supplemented by forecasts based on the authority of experts, or the authority of methods such as econometric modeling.
3. An *expert judgment* is a forecast based on the professional experience and authority of persons who are presumed to have special capabilities to foresee future states of society. Informal judgments are achieved through a process of intuition, which tends to occur quickly, automatically, and effortlessly, some would say on the basis of tacit knowledge.[2]

[1]Forecasts refer to products (forecasts), processes (forecasting), or both.

[2]Daniel Kahneman contrasts System I and System II thinking in "A Perspective on Judgment and Choice: Mapping Bounded Rationality," *American Psychologist* 58, 9 (2003): 697. The term *tacit knowledge* ("we know more than we can say") is usually attributed to Michael Polanyi. See "The Logic of Tacit Inference," *Philosophy* 41, 155 (1966): 1–18 and other works.

By contrast, forecasts based on explicit reasoning tend to be made more slowly, in a more neutral manner, and according to the accepted rules of a scientific discipline or profession. Reasoning (System II), more than intuition (System I), tends to govern extrapolation and prediction (Figure 4.1).

The Value of Forecasting

Policy forecasting, whether based on extrapolation, theory, or expert judgment, provides information about expected policy outcomes. The aims of forecasting are often similar to those of scientific research, which seeks to both understand, predict, and control the human and material environment. Efforts to forecast future societal states are "especially related to control—that is, to the attempt to plan and to set policy so that the best possible course of action might be chosen among the possibilities which the future offers."[3]

Forecasting may contribute to understanding future societal states, thereby achieving some measure of societal guidance. This is frequently accomplished by investigating the effects of past policies and the conditions under which they arise on future policies, an aim suggesting that the future is determined by the past. Although there is a measure of truth in this suggestion, forecasts may also enable policymakers to actively shape the future, despite what may have occurred in the past. The question here is one of identifying the values that can

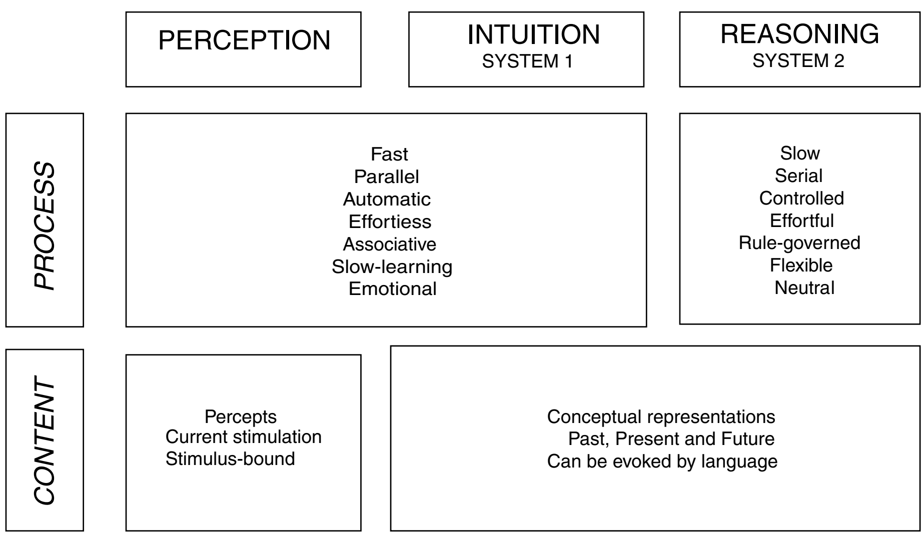


FIGURE 4.1
Processes of Intuition (System I) and Reasoning (System II)

Source: Kahneman (2003): 697.

[3]Irene Taviss, "Futurology and the Problem of Values," *International Social Science Journal* 21, 4 (1969): 574. Harold Lasswell, the founder of the policy sciences, reaches similar conclusions in his discussion of the intelligence function of policymaking. See *Pre-view of the Policy Sciences* (New York: Elsevier, 1971).

and should guide future action. This leads to a second and more important question. Even if the values that guide current actions can be identified, it is questionable whether these values will be retained in the future. Therefore, so-called "guiding predictions" may be incomplete unless they evaluate the desirability of future societal states, an act that would seem to require that we base forecasts on future rather than present preferences. Paradoxically, we must predict future values *before* we can meaningfully predict future states of society.[4]

A concern with future values may complement traditional social science research that seeks predictions based on past trends. Although the past may provisionally determine the future, this will hold true only if thoughtful reflection and analysis do not lead policymakers and citizens to change their values and behavior, or if unpredictable factors and surprises do not intervene to create significant social changes, including those that may be regarded as irreversible processes of chaos and emergent order.[5]

The three types of forecasting, singly and in combination, are nevertheless valuable in reducing some of the uncertainty of attempts to foresee expected policy outcomes, whether of policies are representative of the status quo or of a range of new policies. Armstrong, who has done some of the most thorough, systematic, and practically useful studies of evidence-based forecasting, has reviewed hundreds of forecasts conducted in the last 25 years.[6] Among other results are findings that:

- Predictive forecasts using causal models are approximately three times more accurate than simple extrapolative forecasts. In some cases, these predictive forecasts achieved more than a 70 percent reduction in error, as compared with little reduction in the error of extrapolative forecasts.
- Forecasts based on expert judgment that employed multiple rounds (Delphi technique) improved accuracy in about 70 percent of cases, as compared with methods using one-time expert judgment as part of a largely unstructured process, which is the standard procedure used with panels of experts in areas such as health and medicine, the environment, and the economy.
- Predictive forecasts based on theoretical models to not appear to be more accurate than forecasts based on expert judgment, although forecasts based on theoretical models plus expert judgment outperform extrapolative forecasts.

Limitations of Forecasting

Since the late 1980s, there have been numerous unexpected and counterintuitive political, social, and economic events:

- The abandonment of socialism in the Soviet Union
- The fall of the Berlin Wall

[4]Fred Charles Ikle, "Can Social Predictions Be Evaluated?" *Daedalus* 96 (Summer 1967): 747.

[5]See Alasdair MacIntyre, "Ideology, Social Science, and Revolution," *Comparative Politics* 5, 3 (1973), and Ilya Prigogine and Isabelle Stengers, *Order Out of Chaos* (New York: Bantam Books, 1984).

[6]J. Scott Armstrong, "Findings from Evidence-Based Forecasting: Methods for Reducing Forecast Error." *International Journal of Forecasting* 22, 3 (2006): 583–598. See also Armstrong, *Principles of Forecasting: A Handbook for Researchers and Practitioners* (Norwell, MA: Kluwer Academic Publishers, 2001), which is a critical synthesis of principles applied by forecasters in the last quarter-century.

- Global warming and climate change
- The Arab Spring in the Middle East
- The attack on the World Trade Center and Pentagon on September 11, 2001
- The 2016 election of a populist U.S. President with values and behaviors that are inconsistent with democratic traditions
- An unmanageably huge increase in refugees from Syria and the Middle East

These changes call attention to the importance of forecasting under complex and highly uncertain circumstances that have made Nassim Taleb's "Black Swan" a compelling metaphor of unpredictability.[7] Issues of unpredictability, however, are related to well-documented strengths and limitations of various types of forecasts.[8]

1. *Forecast accuracy.* The accuracy of relatively simple forecasts based on the extrapolation of trend in a single variable, along with relatively complex forecasts based on models incorporating many dozens of variables, has been limited. In 2015, for example, the Obama White House underestimated the 2014 budget increase by $98 billion, or 20.2 percent. A similar record of inaccuracy, defined as the variability of estimates of unemployment, inflation, and economic growth characterizes the estimates of the largest econometric forecasting firms such as Chase Econometrics, Wharton Econometric Forecasting Associates, and Data Resources, Inc. In Ireland, a country experiencing significant unemployment, the average absolute error (ignoring negative and positive values) in forecasting the growth of unemployment in the period 1986–2005 was approximately 4 percent, a rather large error considering the hundreds of thousands of workers affected.[9]
2. *Comparative yield of forecasting methods.* The accuracy of predictions based on complex theoretical models of the economy and of the energy resource system has been roughly comparable to forecasts based on the extrapolation of trend and on expert judgment. If one of the important advantages of simple models is sensitivity to surprising or counterintuitive events, simple models have a comparative advantage over their highly complex counterparts, because the users of highly complex models tend to employ them mechanistically, with little attention to surprising outcomes. As Ascher cautions, if the implications of a forecast are obvious when they are recognized without the aid of a complex model, the value of using one is questionable. If the implications are surprising, the strength of a complex model presumably lies in its ability to identify model assumptions and implications.[10] Yet it is precisely

[7]Nassim Taleb, *The Black Swan: The Impact of the Highly Improbable* (New York: Random House, 2007). The concept of a Black Swan, known to some medieval thinkers as a "rare bird," appears in epistemological writings of philosophers such as Karl Popper. If we accept the proposition that all swans are white, the appearance of a single black swan invalidates the proposition.

[8]A classical work on forecasting is William Ascher, *Forecasting: An Appraisal for Policy Makers and Planners* (Baltimore and London: Johns Hopkins University Press, 1978). Also see Ascher, "The Forecasting Potential of Complex Models," *Policy Sciences* 13 (1981): 247–267; and Robert McNown, "On the Use of Econometric Models: A Guide for Policy Makers," *Policy Sciences* 19 (1986): 360–380.

[9]Economic and Social Research Institute, *Medium-Term Revisions, 2008–2015* (Dublin: May 2008).

[10]Ascher, "The Forecasting Potential of Complex Models," p. 255.

these assumptions and implications—for example, the assumption that a predicted increase in gasoline consumption will not lead to changes in gasoline taxes—that are overlooked or discarded when they are inconsistent with the assumptions of models.

3. *Context.* The assumptions of models and their results are sensitive to institutional, temporal, and historical contexts. Variations in institutional incentive systems are a key aspect of differences in these contexts, as represented by contrasts among government agencies, businesses, and nonprofit research institutes. Forecasting accuracy tends to be greater in university and other nonprofit research institutions than in government agencies or in businesses. In turn, the temporal context of a forecast, as represented by the length of time over which the forecast is made (e.g., 1 year versus 5 years) affects forecast accuracy. The longer the period, the less accurate the forecast. Finally, the historical context of forecasts affects accuracy. The comparatively greater complexity and uncertainty of recent historical periods diminishes forecast accuracy, a pattern that is evident in the growth of forecasting errors over the past 50 years.

Thus, the assumptions that forecasters bring to the process affect the accuracy and comparative yield of forecasts. As Ascher notes, one of the difficulties of assessing the performance of forecasters lies in identifying the assumptions they make. In many forecasts, there is a problem of "assumption drag,"[11] that is, a tendency among forecasters to cling to questionable or plainly implausible assumptions despite contrary evidence. In energy forecasting, for example, the tendency to cling to the assumption that the governments of petroleum-exporting countries (OPEC) will remain stable, despite evidence to the contrary, has decisive effects on forecast accuracy. An important implication of "assumption drag" is that the process of problem structuring discussed in Chapter 3 is an important aspect of forecasting. Indeed, the correction of errors by dissolving or unsolving problems is quite as important in forecasting as it is in any other phase of policy analysis (Figure 4.1).

Societal Futures

Policy forecasts, whether made in the form of extrapolations, predictions, or expert judgments, are used to estimate three types of future societal states: possible futures, plausible futures, and normative futures.[12] *Possible futures* are future societal states that may occur, as distinguished from societal states that eventually do occur. A future state is never certain until it actually occurs, and there are many possible futures. *Plausible futures* (sometimes called *alternative futures*) are future states that, based on evidence or theory, are believed to be likely if policymakers do not intervene to redirect the course of events. By contrast, *normative futures* are societal futures that are consistent with future values. The identification of normative futures narrows the range of potential and plausible futures by linking forecasts to the achievement of values such as economic efficiency, social equity, public enlightenment, wealth, and political power.[13]

---

[11]See Ascher, *Forecasting.*

[12]See David C. Miller, "Methods for Estimating Societal Futures," in *Methodology of Social Impact Assessment,* ed. Kurt Finsterbusch and C. P. Wolf (Stroudsburg, PA: Dowden, Hutchinson & Ross, 1977), pp. 202–210.

[13]Classic analyses of values in public policy are Harold D. Lasswell and Abraham Kaplan, *Power and Society: A Framework for Political Inquiry* (New Brunswick, NJ: Transaction Publishers, 2013); Kurt Baier and Nicholas Rescher,

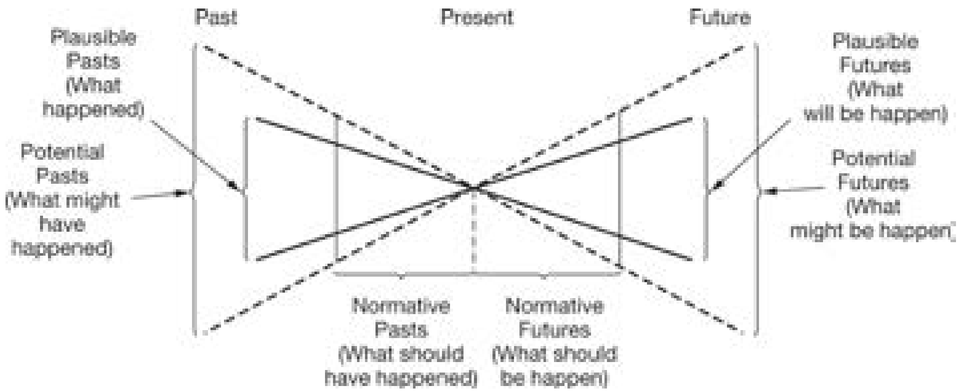


FIGURE 4.2
Three Types of Societal Futures: Potential, Plausible, and Normative

The identification of normative futures not only requires that we deal with values; it also requires that we identify policy alternatives or options that are relevant for their attainment. Although these questions may appear simple, they are in fact difficult. Whose values should the analyst use as a focal point of forecasts? How does an analyst choose among policy alternatives to achieve such values? If analysts use existing policies to identify values and alternatives, they run the risk of simply forecasting a future version of the status quo. If, on the other hand, analysts propose new policy alternatives, they may be charged with bias.

APPROACHES TO FORECASTING

Once goals, objectives, and alternatives have been identified, it is possible to select an approach to forecasting. The analyst must (1) decide what to forecast, that is, determine what the *aim* of the forecast should be; (2) decide how to make the forecast, that is, select the *basis* of the forecast; and (3) choose *methods* that are most appropriate for the aim and basis selected.

Aims and Bases of Forecasts

The *aim* of a forecast is the point of reference of a projection, prediction, or expert judgment. Forecasts have four *aims*:[14]

1. *Forecast the consequences of existing policies.* Forecasts may be used to estimate changes that are likely to occur if no new actions are taken. The status quo (doing

eds., *Values and the Future. The Impact of Technological Change on American Values* (New York, The Free Press, 1969); and Deborah Stone, *Policy Paradox: The Art of Political Decision Making,* rev. ed. (New York: Norton, 2001).
[14]William D. Coplin, *Introduction to the Analysis of Public Policy Issues from a Problem-Solving Perspective* (New York: Learning Resources in International Studies, 1975), p. 21.

nothing) is the policy of taking no new actions. Examples are population projections of the U.S. Bureau of the Census and projections of female labor force participation made by the U.S. Bureau of Labor Statistics.[15]

2. *Forecast the consequences of new policies.* Forecasts may be used to estimate changes in society that are likely if new policies are adopted. In this context, energy demand in 1995 may be based on assumptions about the adoption of new policies to regulate industrial pollution.[16]
3. *Forecast the contents of new policies.* Forecasts may be used to estimate changes in the content of new policies. Some years ago, the Congressional Research Service forecasted the adoption of a 4-week annual paid vacation on the assumption that the government and labor unions will follow the lead of European countries, most of which have adopted 4- or 5-week annual paid vacations for workers.[17]
4. *Forecast the behavior of policy stakeholders.* Forecasts may be used to estimate the probable support and opposition to newly proposed policies. For example, techniques for assessing political feasibility are used to estimate the probability that different stakeholders will support a policy at various stages of the policy process, from adoption to implementation.[18]

The *basis* of a forecast is the set of assumptions or data used to estimate the consequences of existing or new policies, the content of new policies, or the behavior of stakeholders. As we have seen, there are three major bases of forecasts: trend extrapolation, theory, and expert judgment. Each of these bases is associated with one of the three forms of forecasts previously discussed.

Trend extrapolation is the extension into the future of trends observed in the past. Trend extrapolation assumes that what has occurred in the past will also occur in the future, provided that no new policies or unforeseen events intervene to change the course of events. Trend extrapolation is based on *inductive logic,* that is, the process of reasoning from particular observations (e.g., historical time-series data) to general conclusions or claims. In trend extrapolation, we usually start with time-series data, project past trends into the future, and then justify the forecast on the basis of assumptions about regularity and persistence. The logic of trend extrapolation is illustrated in Figure 4.3.

Theoretical assumptions are systematically structured and empirically testable laws and propositions that make predictions about the occurrence of one event on the basis of another. Theoretical assumptions are causal in form, and their role is to explain and predict. The use of theoretical assumptions is based on *deductive logic,* that is, the process of reasoning from general statements, laws, or propositions to particular sets of information and

[15]For example, Howard N. Fullerton Jr., and Paul O. Flaim, "New Labor Force Projections to 1990," *Monthly Labor Review* (December 1976).
[16]See Barry Hughes, *U.S. Energy, Environment and Economic Problems: A Public Policy Simulation* (Chicago, IL: American Political Science Association, 1975).
[17]See Everett M. Kassalow, *Some Labor Futures in the United States*, Congressional Research Service, Congressional Clearing House on the Future, Library of Congress (January 31, 1978).
[18]See Michael K. O'Leary and William D. Coplin, *Everyman's "Prince"* (North Scituate, MA: Duxbury Press, 1976).

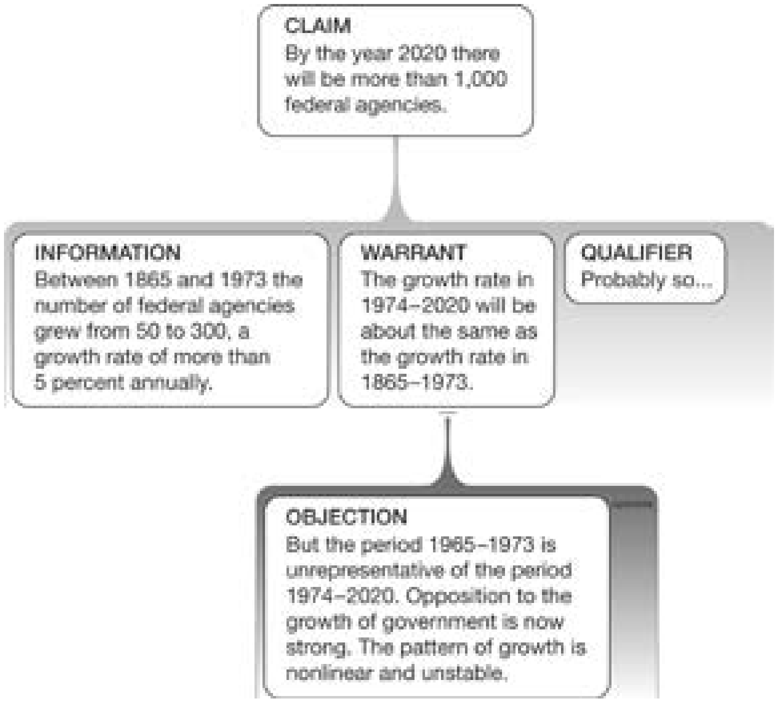


FIGURE 4.3
The Logic of Trend Extrapolation: Inductive Reasoning

claims. For example, the proposition that in "postindustrial" society the predictive knowledge of policy analysts is an increasingly scarce resource that enhances their power may be used to make the claim that the growth of professional analysts in government means that they will have more power than policymakers in coming years (Figure 4.4).

Informed judgment refers to knowledge based on experience and insight, rather than inductive or deductive reasoning. These judgments are usually expressed by experts or knowledgeables and are used when theory and/or empirical data are unavailable or inadequate. Informed judgments are based on *abductive* (*retroductive) logic*,[19] that is, a process of reasoning that begins with claims about the future and then works backwards to the information, assumptions, or theory that supports a claim. A good example of informed judgment as a basis of forecasts is the use of scientists or other knowledgeables to make conjectures about future changes in technology. Through abductive logic, experts may forecast a scenario of automated highways with adaptive automobile autopilots by the year 2050. Experts then work backward to the information, assumptions, and theory required to establish the abductive claim (Figure 4.5).

[19]The preferred term is abduction, as used by Charles Sander Peirce in "Abduction and Induction," *Philosophical Writings of Peirce* (New York: Dover, 1955).

FIGURE 4.4
The Logic of Theoretical Prediction: Deductive Reasoning

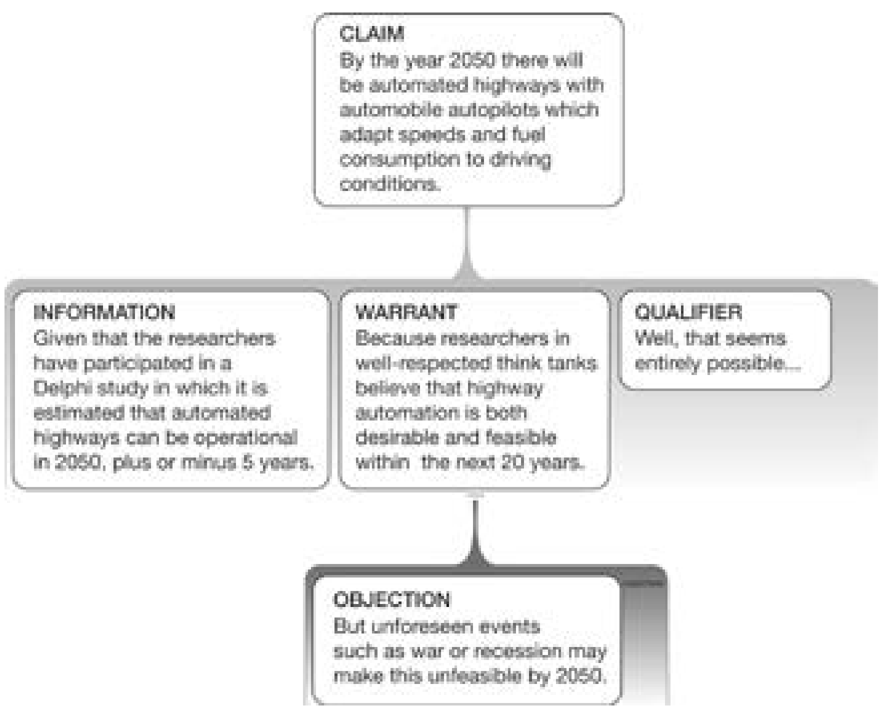


FIGURE 4.5
The Logic of Possible Futures: Abductive Reasoning

Abductive reasoning facilitates the emergence of potential futures out of the present. By contrast, inductive and deductive reasoning are more conservative, because the use of information about past events or the application of established scientific theories restricts consideration of potential (as distinguished from plausible) futures. The restrictive influence of past events and established scientific theories is evident in the forecast of the well-known astronomer, William H. Pickering (1858–1938), who believed that the idea of flying machines carrying large numbers of passengers across the Atlantic is entirely visionary, because "even if a machine could get across with one or two passengers the expense would be prohibitive."[20]

Choosing Methods and Techniques

The choice of an aim and a basis of a forecast helps direct the analyst toward appropriate methods. However, there are literally hundreds of forecasting methods from which to choose.[21] A useful way to simplify the choice of these methods is to group them according to the bases of forecasts discussed earlier. Table 4.1 outlines the three approaches to

TABLE 4.1

Three Approaches to Forecasting

Approach	Basis	Method
Extrapolative Forecasting	Trend Extrapolation	Classical time-series analysis
		Linear trend estimation
		Exponential weighting
		Data transformation
		Catastrophe methodology
Theoretical Forecasting	Theoretical Laws and Propositions	Theory mapping
		Causal modeling
		Regression analysis
		Point and interval estimation
		Correlational analysis
Judgmental Forecasting	Expert Judgment	Classical Delphi
		Policy Delphi
		Cross-impact analysis
		Feasibility forecasting

[20]Quoted in Brownlee Haydon, *The Year 2000* (Santa Monica, CA: RAND Corporation, 1967).

[21]A comprehensive review and evaluation of forecasting methods is Armstrong, *Forecasting Principles: Evidence-Based Forecasting* at www.forecastingprinciples.com. Useful earlier works include Daniel P. Harrison, *Social Forecasting Methodology* (New York: Russell Sage Foundation, 1976); Denis Johnston, "Forecasting Methods in the Social Sciences," *Technological Forecasting and Social Change* 2 (1970): 120–143; Arnold Mitchell and others, *Handbook of Forecasting Techniques* (Fort Belvoir, VA: U.S. Army Engineer Institute for Water Resources, December 1975).

forecasting, their bases, and associated methods. The table serves as an overview of the rest of the chapter.

EXTRAPOLATIVE FORECASTING

Methods of extrapolative forecasting generate forecasts of future societal states based on past, current and historical data. Extrapolative forecasting is often based on *time-series analysis,* which involves the analysis of numerical values collected at multiple points in time and presented sequentially. Time-series analysis provides summary measures of the amount and rate of change in past and future years. Extrapolative forecasting is used to project economic growth, population decline, energy consumption, quality of life, and agency workloads.

Extrapolative forecasting rests on three assumptions:

1. *Persistence.* Trends observed in the past will persist in the future. If energy consumption has increased at the same rate in the past, this secular (or linear) trend will increase in the future.
2. *Regularity.* If observed variations in non-secular patterns have occurred in the past, they will recur in the future. If cycles of economic expansion and contraction have occurred every 7–8 years in the past, they will repeat themselves in the future.
3. *Reliability and validity.* The measurement of secular trends and non-secular patterns is reliable (comparatively accurate) and valid (successful in measuring what it purports to be measuring). For example, crime statistics are relatively accurate measures of criminal offenses when accompanied by victimization surveys, which suggest that crime statistics underestimate criminal offenses by about two-thirds.

When these three assumptions are satisfied, extrapolative forecasting may yield insights into the dynamics of change and a greater understanding of potential future states of society. When any one of these assumptions is violated, extrapolative forecasting methods yield inaccurate or misleading results.

Classical Time-Series Analysis

Classical time-series analysis may be used for extrapolative forecasts. Classical time-series analysis divides time series into four components: secular trend, seasonal variation, cyclical fluctuation, and irregular movement. A *secular trend* is a smooth long-term growth or decline in a time series. Figure 4.6 shows a secular trend in the growth of crimes per 1,000 persons in Chicago over a 30-year period. By convention, the time-series variable is plotted on the *Y*-axis (also called the *ordinate*), and years are plotted on the *X*-axis (also called the *abscissa*). A straight-line trend has been used to summarize the growth of total arrests per 1,000 persons between 1940 and 1970. In other cases, for example, deaths, the straight-line trend shows a long-term decline, while in other cases, for example, the consumption of coal oil, the line graph shows a *curvilinear trend,* that is, a trend in which the numerical values in a time-series display a convex or concave pattern.

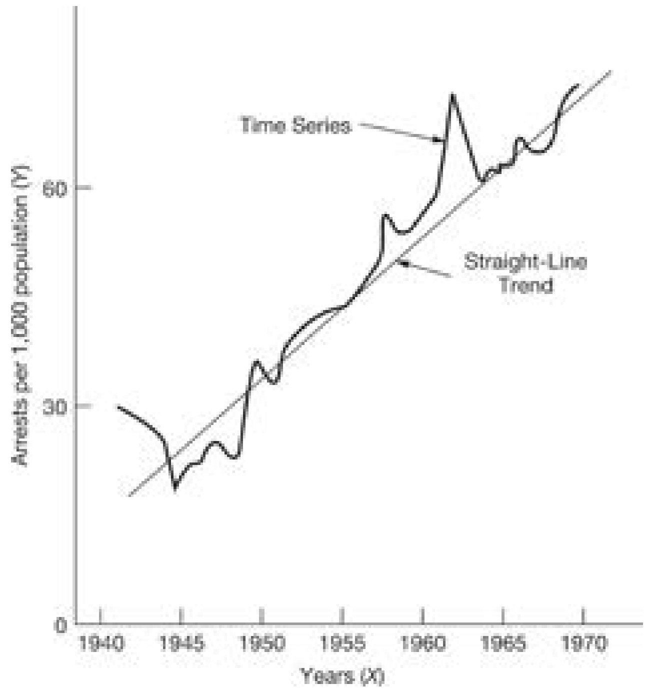


FIGURE 4.6 Demonstration of a Secular Trend: Total Arrests per 1,000 Persons in Chicago, 1940–1970

Source: Adapted from Gurr (1977): 647.

Seasonal variation, as the term suggests, is the variation in a time series that recurs periodically within 1 year or less. The best examples of seasonal variations are the ups and downs of production and sales that follow changes in weather conditions and holidays. The workloads of social welfare, health, and public utilities agencies also frequently display seasonal variations as a result of weather conditions and holidays. For example, the consumption of home heating fuels increases in the winter months and begins to decline in March of each year.

Cyclical fluctuations are also periodic but may extend unpredictably over a number of years. Cycles are frequently difficult to explain because each new cyclical fluctuation may be a consequence of new and unknown factors. Total arrests per 1,000 persons in Chicago, displayed over a period of more than a hundred years, show at least three cyclical fluctuations (Figure 4.7). Each of these cycles is difficult to explain, although the third cyclical fluctuation coincides with the Prohibition Era (1919–1933) and the rise of organized crime. This example points to the importance of carefully selecting an appropriate time frame, because what may appear to be a secular trend may in fact be part of some larger long-term pattern of cyclical fluctuations. Note also that total arrests per 1,000 persons were higher in the 1870s and 1890s than in most years from 1955 to 1970.

The interpretation of cyclical fluctuations frequently must account for the presence of *irregular movements*, that is, unpredictable variations in a time series that appear to follow no regular pattern. Irregular movements may be the result of many factors (e.g., changes in government, strikes, natural disasters). As long as these factors are unaccounted for, they

FIGURE 4.7
Demonstration of Cyclical Fluctuations: Total Arrests per 1,000 Persons in Chicago, 1868–1970

Source: Gurr (1977): 647.

are treated as random error, that is, unknown sources of variation that cannot be explained in terms of secular trends, seasonal variations, or cyclical fluctuations. For example, the irregular movement in total arrests that occurred in 1957 (see Figure 4.6) might be regarded as an unpredictable variation in the time series that follows no regular pattern. On closer inspection, however, the sharp temporary upswing in arrests is explained by the changes in record keeping that came into effect when Orlando Wilson was appointed chief of police.[22] This example points to the importance of understanding the historical and institutional context of changes in time series.

Linear Trend Estimation

A standard technique for extrapolating trends is *linear trend estimation,* a procedure that uses the method of regression analysis to obtain statistically reliable estimates. Linear regression is based on assumptions of persistence, regularity, and measurement reliability. When linear regression is used to estimate trend, it is essential that observed values in a time series are not curvilinear, because any significant departure from linearity will produce an unreliable forecast. Nevertheless, linear regression may be used to remove the linear trend component from a series that displays seasonal variations or cyclical fluctuations.

[22]See Donald T. Campbell, "Reforms as Experiments," *American Psychologist* 24, 4 (1969): pp. 409–429.

Regression analysis has two important properties:

1. *Deviations cancel.* The sum of the differences between observed values in a time series and values lying along a straight-line trend (called a *regression line*) will always equal zero. Thus, if the trend value (Y_t) is subtracted from its corresponding observed value (Y) for all years in a time series, the total of these differences (called *deviations*) will equal zero. When the observed value for a given year lies *below* the regression line, the deviation ($Y - Y_t$) is *negative.* When the observed value is *above* the regression line, the deviation ($Y - Y_t$) is *positive.* These negative and positive values cancel, such that $\Sigma(Y - Y_t) = 0$.
2. *Squared deviations are a minimum.* If we square each deviation by multiplying each deviation value by itself and add them all up, the sum of these squared deviations will always be a minimum or least value. This means that linear regression minimizes the distances between the regression line and all observed values of Y in the series. In other words, regression analysis is the most efficient way to draw a trend line through a series of observed data points.

These two properties of regression analysis are illustrated with hypothetical data in Figure 4.8.

The computation of linear trend with regression analysis is illustrated in Table 4.2 with data on energy consumption. In column (3), years in the time series have been coded by calculating the numerical distance (x) of each year from 1973, which is the middle of the series. The middle of the series is given a coded value of zero and treated in this case as the

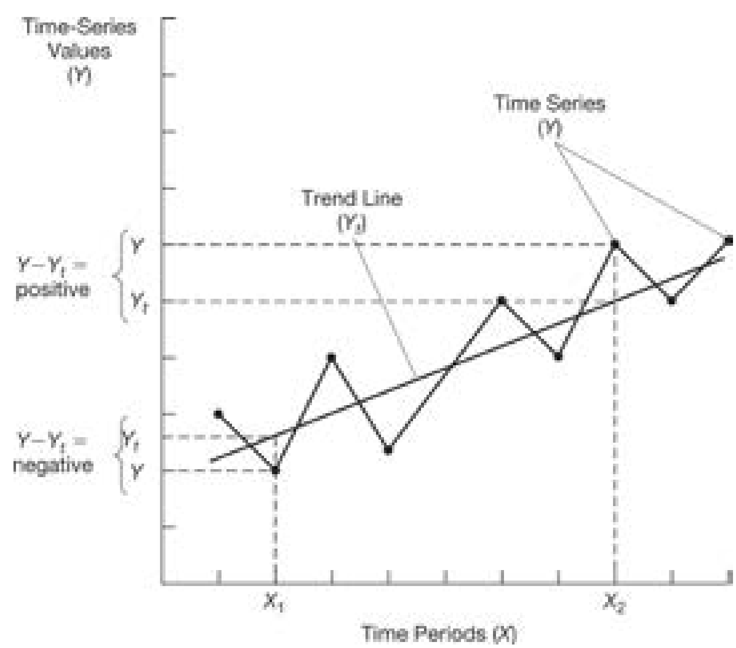


FIGURE 4.8 Two Properties of Linear Regression

Note: $\Sigma(Y - Y_t) = 0$; $\Sigma(Y - Y_t)^2$ = a minimum (least) value.

TABLE 4.2

Time-Series Data on Total Energy Consumption Used in Linear Regression

Years	Energy Consumption	Coded Time Value	Columns (2) by (3)	Column (3) Squared	Trend Value for Energy Consumption
(X)	(Y)	(x)	(xY)	(x^2)	(Y_t)
(1)	(2)	(3)	(4)	(5)	(6)
1970	66.9	−3	−200.7	9	68.35
1971	68.3	−2	−136.6	4	69.31
1972	71.6	−1	−71.6	1	70.27
1973	74.6	0	0	0	71.24
1974	72.7	1	72.7	1	72.20
1975	70.6	2	141.2	4	73.17
1976	74.0	3	222.0	9	74.13
$n = 7$	$\Sigma Y = 498.7$	$\Sigma x = 0$	$\Sigma(xY) = 27.0$	$\Sigma(x^2) = 28$	$\Sigma Y_t = 498.7$

$$Y_t = a + b(x)$$

$$a = \frac{\Sigma Y}{n} = \frac{498.7}{7} = 71.243$$

$$b = \frac{\Sigma(xY)}{\Sigma(x^2)} = \frac{27.0}{28.0} = 0.964$$

$$Y_t = 71.24 + 0.964(x)$$

origin of values of x to the right and the left of zero.[23] The values of x range from −3 to +3 and are deviations from the origin. Observe also the computational procedures used in columns (4) and (5) of Table 4.2. The value of xY is calculated by multiplying values of energy consumption (Y) by each coded time value (x), as shown in column (4). The value of x^2, shown in column (5), is calculated by multiplying each value of x by itself, that is, squaring all the x values in column (3). Finally, column (6) contains trend values of the time-series variable (Y_t). These trend values, which form a straight line, are calculated according to the following equation:

$$Y_t = a + b(x)$$

[23]For purposes of simplicity, the origin of the series has been set at zero and intuitively understandable statistical notation has been used. Standard methods and statistical notation used in regression analysis are presented in texts on econometrics including Jeffrey M. Wooldridge, *Introductory Econometrics: A Modern Approach* (Boston, MA: Nelson Education, 2015); James H. Stock and Mark W. Watson, *Introduction to Econometrics* (Boston, MA: Addison Wesley, 2003); and Peter Kennedy, *A Guide to Econometrics* (Cambridge, MA: MIT Press, 2003).

where

Y_t = the trend value for a given year
a = the value of Y_t when $x = 0$
b = the slope of the trend line representing the change in Y_t for each unit of time
x = the coded time value for any year, as determined by its distance from the origin.

Once the values of a and b have been calculated, estimates of total energy consumption may be made for any year in the *observed* time series or for any (unobserved) *future* year. For example, Table 4.2 shows that the trend value for energy consumption in 1972 is 70.27 quadrillion BTUs. To project total energy consumption for 1980, we set the value of x at 7 (i.e., seven time periods away from 1973, the origin) and solve the equation $Y_{t(1980)} = a + b(x)$. The formula for computing the value of a is

$$a = \frac{\Sigma Y}{n}$$

where

ΣY = the sum of observed values in the series
n = the number of years in the observed time series

The formula for computing the value of b is

$$b = \frac{\Sigma(xY)}{\Sigma\left(x^2\right)}$$

where

(ΣxY) = the sum of the products of the coded time values and the observed values in the series [see column (4) of Table 4.2]
$\Sigma\ (x^2)$ = the sum of the squared coded time values [see column (5) of Table 4.2]

The calculations in Table 4.2 not only permit us to estimate trend values for energy consumption in the observed series; they also enable us to project total energy consumption for any given year in the future. Thus, the estimate of total energy consumption in 1980 [$Y_{t(1980)}$] is 77.99 quadrillion BTUs.

In illustrating the application of linear regression, we have used a time series with an odd number of years. We treated the middle of the series as the origin and coded it as zero. Obviously, there are many time series that contain an even number of years, and a different procedure is used for determining each coded time value (x), because there is no middle year. The procedure used with even-numbered time series is to divide the series into two equal parts and code the time values in intervals of two (rather than in intervals of one, as in odd-numbered series). Each year in Table 4.3 is exactly two units away from its neighbor, and the highest and lowest coded time values are +7 and –7, rather than +3 and –3. Note that the size of the interval need not be 2; it could be any number, provided that each year is equidistant from the next one in the series. Increasing the size of the interval will not affect the computation of results.

Despite its precision in extrapolating secular trends, linear regression is limited by several conditions. First, the time series must be *linear,* that is, display a constant increase or decrease in values along the trend line. If the pattern of observations is *nonlinear* (i.e., the amount of change increases or decreases from one time period to the next), other techniques must be used. Some of these techniques require fitting various types of curves to nonlinear time series. Second, plausible arguments must be offered to show that historical patterns will *persist* in the future, that is, continue in much the same form in subsequent years as they have in past ones. Third, patterns must be *regular,* that is, display no sharp discontinuities. Unless all of these conditions are present, linear regression should not be used to extrapolate trends.

The Statistical Package for the Social Sciences (SPSS) output for the regression analyses computed by hand in Table 4.2 and 4.3 are displayed in Exhibit 4.1. As may be seen by comparing Table 4.2 with the SPSS output, the results are identical. The SPSS output

TABLE 4.3

Linear Regression with an Even-Numbered Series

Years (X) (1)	Energy Consumption (Y) (2)	Coded Time Value (x) (3)	Columns (2) By (3) (xY) (4)	Column (3) Squared (x^2) (5)	Trend Value for Energy Consumption (Y_t) (6)
1969	64.4	−7	−450.8	49	66.14
1970	66.9	−5	−334.5	25	67.36
1971	68.3	−3	−204.9	9	68.57
1972	71.6	−1	−71.6	1	69.78
1973	74.6	1	74.6	1	71.00
1974	72.7	3	218.1	9	72.21
1975	70.6	5	353.0	25	73.43
1976	74.0	7	518.0	49	74.64
$n = 8$	$\Sigma Y = 563.1$	$\Sigma x = 0$	$\Sigma\ (xY) = 101.9$	$\Sigma\ (x^2) = 168$	$\Sigma Y_t = 563.1$

$$Y_t = a + b(x)$$

$$a = \frac{\Sigma Y}{n} = \frac{563.1}{8} = 70.39$$

$$b = \frac{\Sigma(xY)}{\Sigma\left(x^2\right)} = \frac{101.9}{168} = 0.607$$

$$Y_t = 70.39 + 0.607(x)$$

$$Y_{t(1980)} = 70.39 + 0.607(15) = 79.495 \text{ quadrillion BTUs}$$

identifies the dependent variable (which was named ENCONS for energy consumption); the number of observations in the sample (n = 7), the correlation coefficient expressing the strength of the relationship between time and energy consumption (R: 0.729); and the squared correlation coefficient (SQUARED MULTIPLE R: 0.531), which expresses the proportion of variance in the dependent variable (ENCONS) explained by the independent variable (TIMECODE).

EXHIBIT 4.1

SPSS Output for Tables 4.2 and 4.3

Model Summary[b]

Model	*R*	*R* Square	Adjusted *R* Square	Standard Error of the Estimate	Durbin Watson
1	.729[a]	.531	.437	2.14575994	1.447

[a]Predictors: (Constant), TIMECODE
[b]Dependent Variable: ENCONS

ANOVA[b]

Model		Sum of Squares	*df*	Mean Square	*F*	Significance
1	Regression	26.036	1	26.036	5.655	.063[a]
	Residual	23.021	5	4.604		
	Total	49.057	6			

[a]Predictors: (Constant), TIMECODE
[b]Dependent Variable: ENCONS

Coefficients[a]

		Unstandardized Coefficients		Standardized Coefficients		
Mode		*B*	Std. Error	Beta	*t*	Sig.
1	(Constant)	71.243	.811		87.843	.000
	TIMECODE	.964	.406	.729	2.378	.063

[a]Dependent Variable: ENCONS

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	*N*
Predicted Value	68.349998	74.135712	71.242857	2.08309522	7
Residual	−2.571429	3.3571429	8.120E−15	1.95880187	7
Std. Predicted Value	−1.389	1.389	.000	1.000	7
Std. Residual	−1.198	1.565	.000	.913	7

[a]Dependent Variable: ENCONS

For present purposes, the most important part of the SPSS output gives the coefficient for the constant, *a*, which is 71.243—this is the value of *Y*, energy consumption, when *X* equals zero. Line 5 gives the coefficient for the slope of the regression line, *b*, which is 0.964—this is the amount of change in *Y*, energy consumption (ENCONS) for each unit change in *X*. This part of the output enables us to write the regression equation and, using the coefficients in this equation, forecast the value of energy consumption for any future year. The regression equation for Table 4.2 is

$$Y_t = 71.243 + 0.964(X)$$

The regression equation for Table 4.3, which has eight rather than 7 years in the time series, is

$$Y_t = 70.388 + 0.607(X)$$

Note how much of a difference there is in the slopes after adding only 1 year to the series.

Many time-series data of concern to policy analysts—for example, data on crime, pollution, public expenditures, urbanization—are nonlinear. A variety of techniques has been developed to fit nonlinear curves to patterns of change that do not display a constant increase or decrease in the values of an observed time series. Although these techniques are not presented in detail, we discuss some of their properties and underlying assumptions. Readers who wish to investigate these forecasting techniques in greater depth should consult advanced texts on time-series analysis.[24]

Nonlinear Time Series

Time series that do not meet conditions of linearity, persistence, and regularity fall into five main classes (Figure 4.9):

1. *Oscillations.* Here, there are departures from linearity, but only *within* years, quarters, months, or days. Oscillations may be persistent and regular (e.g., most police arrests occur between 11 P.M. and 2 P.M. throughout the year) but not show a constant increase or decrease within the period under examination (Figure 4.9[a]). Oscillations *within* years may occur in conjunction with long-term secular trends *between* years. Examples include seasonal variations in unemployment, monthly variations in agency workloads, and daily variations in levels of pollutants.
2. *Cycles.* Cycles are nonlinear fluctuations that occur between years or longer periods of time. Cycles may be unpredictable or occur with persistence and regularity. Whereas the overall pattern of a cycle is always nonlinear, segments of a given cycle may be linear or curvilinear (Figure 4.9[b]). Examples are business cycles and the "life cycles" of academic fields, scientific publications, and civilizations.
3. *Growth curves.* Departures from linearity occur *between* years, decades, or some other unit of time. Growth curves evidence cumulative *increases* in the rate of growth

[24]See G.E.P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control* (San Francisco, CA: Holden-Day, 1969); Spyros Makridakis, Steven C. Wheelwright, and Rob J. Hyndman, *Forecasting Methods and Applications* (New York: John Wiley & Sons, 2008).

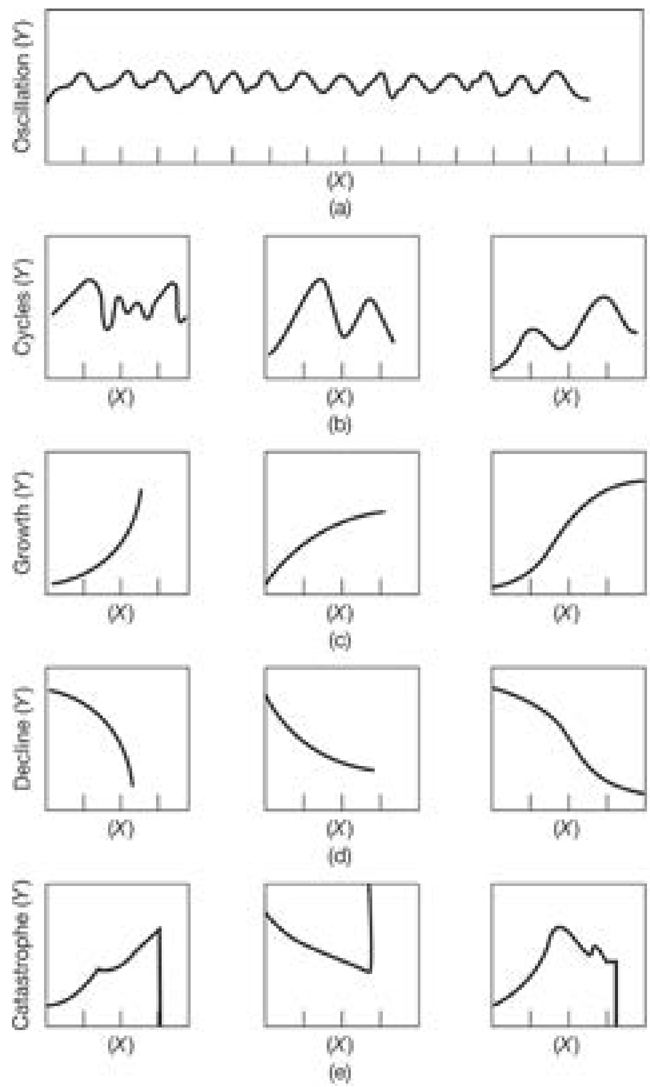


FIGURE 4.9
Five Classes of Nonlinear Time Series

in a time series, cumulative *decreases* in this rate of growth, or some *combination* of the two (Figure 4.9[c]). In the latter case, growth curves are S-shaped and are called *sigmoid* or *logistic* curves. Growth curves, which developed out of studies of biological organisms, have been used to forecast the growth of industry, urban areas, population, technology, and science. Although growth curves are not linear, they are nevertheless persistent and regular.

4. *Decline curves.* Here, departures from linearity again occur *between* years, decades, or longer periods. In effect, decline curves are the counterpart of growth curves. Decline curves evidence either cumulative increases or decreases in the rate of decline in a time series (Figure 4.9[d]). Increasing and decreasing rates of decline may be

combined to form curves with different shapes. Patterns of decline are sometimes used as a basis for various dynamic or life cycle perspectives of the decline of civilizations, societies, and urban areas. Decline curves are nonlinear but regular and persistent.

5. *Catastrophes.* The main characteristic of time-series data that are catastrophic is that they display sudden and sharp discontinuities. The analysis of catastrophic change, a field of study founded by the French mathematician René Thom, not only involves nonlinear changes over time but also patterns of change that are discontinuous (Figure 4.9[e]). Examples include sudden shifts in government policy during war (attack or withdrawal), the collapse of stock exchanges in times of economic crisis, and the sudden change in the density of a liquid as it boils.[25]

The growth and decline curves illustrated in Figures 4.9(c) and (d) cannot be described by a straight-line trend. Patterns of growth and decline are not cyclical, and the trend is best described as *exponential growth* or *decline,* that is, growth or decline whereby the values of a quantity increase or decrease at an increasing or decreasing rate. The growth of federal government organizations between 1789 and 1973 (Figure 4.10) is an example of exponential

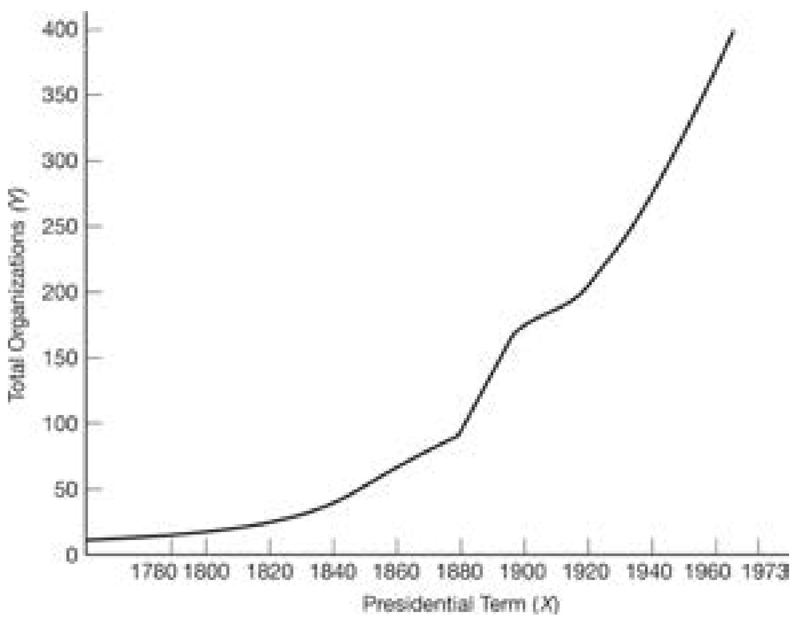


FIGURE 4.10
Growth of Federal Government Organizations in the United States by Presidential Term, 1789–1973

Source: Kaufman (1976): 62.

[25]See C. A. Isnard and E. C. Zeeman, "Some Models from Catastrophe Theory in the Social Sciences," in *The Use of Models in the Social Sciences,* ed. Lyndhurst Collins (Boulder, CO: Westview Press, 1976), pp. 44–100.

growth. After 1789, the total number of organizations began to grow slowly, until about 1860 when the growth rate began to accelerate. Clearly, this growth trend is very different from the secular trends and cyclical variations examined so far.[26]

Techniques for fitting curves to processes of growth and decline are more complex than those used to estimate secular trend. Although many of these techniques can be adapted to the requirements of linear regression, the time-series variable (Y) must be transformed. Some of these transformations involve roots $\left(\sqrt[n]{Y}\right)$, whereas others require logarithms ($\log Y$) or exponents (e^x). In each case, the aim is to express mathematically changes in a time series that increase by increasing or decreasing amounts (or, conversely, decrease by increasing or decreasing amounts). These techniques will not be presented here, but their logic is sufficiently important for public policy analysis to warrant further investigation.

A simple illustration will serve best to clarify techniques for curve fitting. Recall the model of compound interest. This model states that

$$S_n = (1+r)^n\ S_0$$

where

S_n = the amount to which a given investment will accumulate in a given (n) number of years
S_0 = the initial amount of the investment
$(1 + r)^n$ = a constant return on investment (1.0) plus the rate of interest (r) in a given (n) number of years

Imagine that the manager of a small municipality of 10,000 persons is faced by the following situation. In order to meet immediate out-of-pocket expenses for emergencies, it was decided in 1970 that a sum of $1,000 should be set aside in a special checking account. The checking account is immediately accessible but earns no interest. In 1971 and subsequent years, because of rising inflation, the manager began to increase the sum in the special account by $100 per year. The increase in funds, illustrated in Figure 4.11(a), is a good example of the kind of linear trend we have been examining. The time series increases by constant amounts ($100 per year). By 1980, there was $2,000 in the special account. In this case, the time-series values (Y) are identical to the trend values ($Y_t = Y$), because all values of Y are on the trend line.

Now consider what would have occurred if the manager had placed the funds in an interest-bearing account (we will assume that there are no legal restrictions and that withdrawals can be made without penalties). Assume that the annual rate of interest is 10 percent compounded annually. Assume also that only the original $1,000 was left in the account for the 10-year period, that is, no further deposits were made. The growth of city funds, illustrated in Figure 4.11(b), is a good example of the kind of growth trend we

[26]The outstanding work on processes of growth in science and other areas is Derek de Sólla Price, *Little Science, Big Science* (New York: Columbia University Press, 1963).

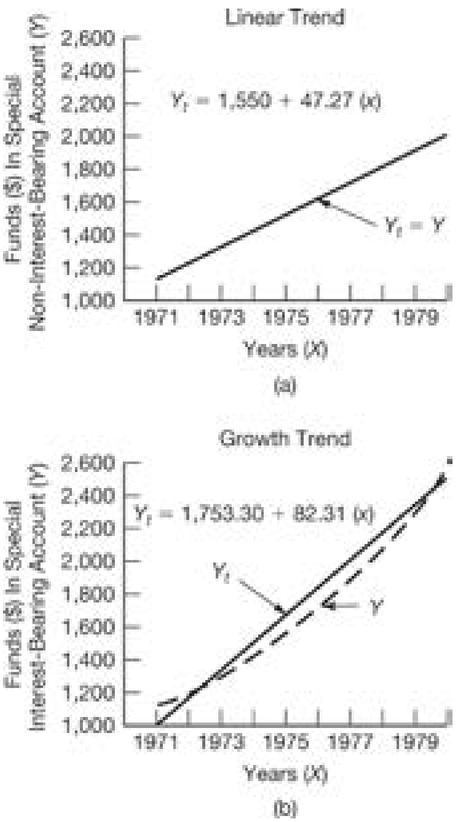


FIGURE 4.11
Linear versus Growth Trends

have been discussing. City funds increased by increasing amounts over the 10-year period (but at a constant interest rate compounded annually), and the total funds available at the end of 1980 was \$2,594, compared with \$2,000 in the non-interest-bearing account (Figure 4.11[a]). In the first case (no interest), it takes 10 years to double the original amount. No accumulation above the original \$1,000 comes from interest. In the second case, it takes a little more than 7 years to double the original amount, and all of the additional (\$1,594) accumulation comes from interest. The values used to calculate accumulation at the end of 1980 are

$$
\begin{aligned}
S_n &= (1+r)^n \, S_0 \\
&= (1+r)^{10} (\$1{,}000) = (2.5937)(\$1{,}000) \\
S_{10} &= \$2{,}594
\end{aligned}
$$

Note that accumulation for any given year (e.g., 1975 as the 5th year) may be calculated simply by substituting the appropriate values in the formula $S_5 = (1 + r)^5(\$1{,}000) = (1.6105)(\$1{,}000) = \$1{,}610.51$.

Consider now the limitations of linear regression in estimating nonlinear growth trend such as that illustrated in Figure 4.11(b). The linear regression equation of Figure 4.11(b) [Y_t = 1,753.30 + 82.31(x)] will produce an inaccurate forecast. For example, a 1990 forecast using the linear regression equation yields a trend estimate of \$4,140.29 [$Y_{t(1990)}$ = 1,753.30 + 82.31(29) = \$4,140.29]. By contrast, the compound interest formula, which exactly represents the nonlinear growth in accumulated funds, produces a 1990 estimate of \$6,727.47 [$S_{20} = (1.1)^{20}$ (\$1,000) = \$6,727.47]. The linear regression estimate is therefore highly inaccurate.

Fortunately, the linear regression technique may be adapted to estimate growth trends. For regression to do this, however, it is necessary to change our original linear equation [$Y_t = a + b(x)$] to a nonlinear one. Although there are many ways to do this, two of the most common procedures are exponential weighting and data transformation.

Exponential Weighting

In *exponential weighting,* the analyst may raise one of the terms in the regression equation to a power, for example, square the value of (x), or add another term, also raised to a power, to the equation [e.g., $c(x^2)$]. The more pronounced the increase (or decrease) in the observed growth pattern, the higher the power necessary to represent it. For example, if we have slowly increasing amounts of change, such as those illustrated by the compound interest formula (Figure 4.11[b]), we might add to the original linear regression equation a third term [$c(x^2)$] that has been raised to a power by squaring the coded time value (x). The equation (called a *second-degree parabola*) then reads

$$\begin{aligned} Y_t &= a + b(x) + c\left(x^2\right) \\ &= 1753.30 + 82.31(x) + 1.8\left(x^2\right) \end{aligned}$$

The only part of the equation that is not linear is the squared time value (x^2). This means that any value of x computed for given years will increase by a power of 2 (e.g., coded values of −5 and +9 become 25 and 81, respectively). The higher the original x value, the greater the amount of change, because the squares of larger numbers produce disproportionately larger products than do the squares of smaller numbers. Whereas a coded time value (x) of 9 is 3 times the value of 3 in the linear regression equation, in the second-degree parabola the same year is nine times the value of the corresponding year:

$$\frac{9}{3} = 3 \text{ and } \frac{9^2}{3^2} = 9$$

It is now easier to visualize what it means to say that growth processes exhibit increasing increases in a time series.

Data Transformation

A second procedure used to adapt linear regression techniques to processes of growth and decline is *data transformation.* Whereas exponential weighting involves the writing of a new and explicitly nonlinear regression equation [e.g., $Y_t = a + b(x) + c(x^2)$], the transformation of data permits analysts to work with the simple linear equation [$Y_t = a + b(x)$], but only after values of the time-series variable (Y) have been appropriately transformed. One way to transform the time-series variable is to take its square root $\left(\sqrt{Y}\right)$. Another way to transform data is to take the natural *logarithm* of values of the time-series variable. The common (*base 10*) logarithm of a number is the power to which 10 must be raised to produce that number. For example, 2 is the logarithm of 100, because 10 must be raised to the power of 2 to produce 100 ($10 \times 10 = 10^2 = 100$). The natural (*base e*) logarithm of a number is the power to which the base *e*, which equals 2.71828, must be raised to produce that number. For example, the natural logarithm of 100 is 4.6052, because 2.71828 raised to the power 4.6052 equals 100. The abbreviation for base 10 logarithms is *log*, and the abbreviation for base *e* logarithms is *ln*. For readers with no experience with logarithms and roots, it is easier to grasp the nature of these transformations if we study a hypothetical time series that exhibits rapid growth (Table 4.4). We can then observe the consequences of taking roots and logarithms of the time-series variable before solving the simple linear regression equation.

Note that all computations used in Table 4.4 are identical to those already used to estimate trend by using linear regression. The only difference is that values of Y have been transformed, either by taking their square roots or by taking their common (*base 10*) logarithms. However, the linear regression procedure is inappropriate for describing the sharp growth trend that increases by multiples of 10 in each period [column (2), Table 4.4]. The linear equation [$Y_t = a + b(x)$] produces a 1981 forecast of 85,227, when the actual value of the series will be 1,000,000! (We are assuming, of course, that the series will grow by a factor of 10 in 1981 and subsequent years.) The linear equation based on the square root transformation $\left[\sqrt{Y_t} = a + b(x)\right]$ helps very little with this explosive time series. The transformation results in a 1981 estimate of 94,249, hardly much of an improvement.

Only the logarithmically transformed linear equation [$\log Y = a + b(x)$] gives us the exact value for 1981. In effect, what we have done is "straighten out" the growth trend by taking the common logarithm of Y, thus permitting the use of linear regression to extrapolate trend. This does *not* mean that the trend is actually linear (obviously, it is not). It only means that we can use linear regression to make a precise forecast of the value of the *nonlinear* trend in 1981 or any other year. However, to make a trend estimate for 1981, we must convert the logarithmic value (6.0) back into the original value of the Y variable. This is done (in Table 4.4) by taking the *antilogarithm* of 6.0 (abbreviated "*antilog* 6.0"), which is 1,000,000 (an antilogarithm is the number corresponding to the logarithm, e.g., *antilog* 2 = 100, *antilog* 3 = 1,000, *antilog* 4 = 10,000, and so on). Note also that roots must also be reconverted by raising the trend estimate by the appropriate power (see Table 4.4).

Now that we understand some of the fundamentals of estimating growth trends, let us return to our example of the manager of the small municipality and the problem of compound interest. We know from the compound interest model that \$1,000 invested at an interest rate of 10 percent (compounded annually) will accumulate to \$6,727.47 by the end of 1990. That is,

$$\begin{aligned} S_n &= (1+r)^n \; S_0 \\ &= (1+0.10)^{20} \; (\$1,000) \\ S_{20} &= \$6,727.50 \end{aligned}$$

TABLE 4.4

Square Root and Logarithm of a Time Series Exhibiting Rapid Growth

Year	Original Value of Time-Series Variable	Square Root of Time-Series Variable	Logarithm of Time-Series Variable	Coded Time Value	Time Value Squared	Columns (2) by (5)	Columns (3) by (5)	Columns (4) by (5)	Trend Estimates		
(X) (1)	(Y) (2)	$(\sqrt{Y})_t$ (3)	(log Y) (4)	(x) (5)	(x^2) (6)	(xY) (7)	$(x\sqrt{Y})_t$ (8)	(x log Y) (9)	Y_t (10)	$\sqrt{Y}_t$ (11)	log Y_t (12)
1976	10	3.16	1.0	−2	4	−20	−6.32	−2	−19,718	−51.0	1.0
1977	100	10.00	2.0	−1	1	−100	−10.0	−2	1,271	20.6	2.0
1978	1,000	31.62	3.0	0	0	0	0	0	22,260	92.2	3.0
1979	10,000	100.00	4.0	1	1	10,000	100.00	4	43,249	163.8	4.0
1000	100,000	316.23	5.0	2	4	200,000	632.46	10	64,238	235.4	5.0
5	111,300	461.01	15.0	0	10	209,880	716.14	10.0	11,300	461.01	15.0

$$Y_t = a + b(x)$$
$$a = \frac{\Sigma Y}{n} = \frac{111{,}300}{5} = 22{,}260$$
$$b = \frac{\Sigma(xY)}{\Sigma\left(x^2\right)} = \frac{209{,}880}{10} = 20{,}988$$
$$Y_t = 22{,}260 + 20{,}988(x)$$
$$Y_{t(1981)} = 22{,}260 + 20{,}988(3)$$
$$= 85{,}224$$

$$\sqrt{Y_t} = a + b(x)$$
$$a = \frac{\Sigma\sqrt{Y}}{n} = \frac{461.01}{5} = 92.2$$
$$b = \frac{\Sigma\left(x\sqrt{Y}\right)}{\Sigma\left(x^2\right)} = \frac{716.14}{10} = 71.6$$
$$\sqrt{Y_t} = 92.2 + 71.6(x)$$
$$Y_{t(1981)} = \left[92.2 + 71.6(3)\right]^2$$
$$= 94{,}249$$

$$\log Y = a + b(x)$$
$$a = \frac{\Sigma \log Y}{n} = \frac{15}{5} = 3.0$$
$$b = \frac{\Sigma(x \log Y)}{\Sigma\left(x^2\right)} = \frac{10}{10} = 1.0$$
$$\log Y = 3.0 + 1.0(x)$$
$$Y_{t(1981)} = \text{antilog } 3.0 + 1.0(3)$$
$$= \text{antilog } 6.0$$
$$= 1{,}000{,}000$$

If we use the linear regression equation to extrapolate to 1990, we will obtain \$4,140.29. This is \$2,587.18 in error (\$6,727.47–4,140.29 = \$2,587.18). But what can we do about this sizable error, given that we would like to use the linear regression model to forecast the growth trend? The easiest way to deal with this problem is to find the base 10 logarithms of all values of *Y* and use these with the linear regression equation.

This has been done in Table 4.5, which shows that the 1990 trend estimate based on the logarithmic transformation should be $\log Y_{t(1990)} = 3.828$. When we take the antilogarithm of this value, we find that it equals approximately \$6,730, an estimate that is very close to that produced by the compound interest formula (note: small errors are due to rounding). Hence, through a logarithmic transformation of the time-series variable, we have applied linear regression to a nonlinear process. The same kinds of logarithmic transformations are regularly used in public policy analysis to extrapolate other nonlinear trends, including national income, population, and government expenditures.

Catastrophe Methodology

No matter how successfully we use linear regression to forecast growth and decline, there is one condition that must be present: The processes we seek to understand and extrapolate must be relatively smooth and continuous. Yet as we have seen (Figure 4.9[e]), many time series are discontinuous. It is here where catastrophe methodology, a field within the special branch of mathematics called topography, provides useful insights. *Catastrophe methodology,* which involves the systematic study and mathematical representation of discontinuous processes, is specifically designed to forecast trends when small changes in one variable (e.g., time) produce large changes in another variable. The point at which a small change in one variable results in a dramatic shift in another variable is called a *catastrophe,* a term that refers to a mathematical theorem that classifies discontinuous processes into five major types, rather than to any sense of impending doom or disaster. According to its founder, René Thom, it is a methodology for studying elementary types of discontinuity in nature and society.[27]

Catastrophe methodology is much too complex to survey here. We should nevertheless be aware of some of its major assumptions and applications.

1. *Discontinuous processes.* Many of the most important physical, biological, and social processes are not only curvilinear; they are also abrupt and discontinuous. An example of discontinuous processes in the social realm is the sudden shifts in public attitudes that sometimes follow smooth and gradually evolving changes in public opinion.
2. *Systems as wholes.* Social systems as a whole frequently exhibit changes that are not the simple sum of their parts. Sudden shifts may occur in the structure of a social system as a whole, although the system's parts may have changed gradually and

[27]See Isnard and Zeeman, "Some Models from Catastrophe Theory in the Social Sciences." René Thom's major work is *Stabilité structurelle et morphogenèse,* trans. D. Fowler (New York: W. A. Benjamin, 1975). Related to catastrophe methodology is work on chaos theory, which offers explanations of discontinuous change. See Prigogine and Stengers, *Order Out of Chaos.*

TABLE 4.5

Linear Regression with Logarithmic Transformation of Time Series

Years	Time-Series Variable	Logarithm of Time-Series Variable	Coded Time Value	Squared Time Value	Columns (3) By (4)	Trend Estimate
(X)	(Y)	(log Y)	(x)	(x^2)	(x log Y)	(log Y_t)
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1971	1100	3.041	−9	81	−27.369	3.0413
1972	1210	3.083	−7	49	−21.581	3.0827
1973	1331	3.124	−5	25	−15.620	3.1241
1974	1464	3.165	−3	9	−9.495	3.1655
1975	1611	3.207	−1	1	−3.207	3.2069
1976	1772	3.248	1	1	3.248	3.2483
1977	1949	3.290	3	9	9.870	3.2897
1978	2144	3.331	5	25	16.655	3.3311
1979	2358	3.373	7	49	23.611	3.3725
1980	2594	3.414	9	81	30.726	3.4139
10	17533	32.276	0	330	6.838	32.2760

$$\log Y_t = a + b(x)$$

$$a = \frac{\Sigma \log Y}{n} = \frac{32.276}{10} = 3.2276$$

$$b = \frac{\Sigma(x \log Y)}{\Sigma\left(x^2\right)} = \frac{6.838}{330} = 0.0207$$

$$\log Y_t = 3.2276 + 0.0207(x)$$

$$\log Y_{t(1990)} = 3.2276 + 0.0207(29) = 3.828$$

$$Y_{t(1990)} = \text{antilog } 3.828 = \$6{,}745.27$$

smoothly. For example, public opinion on major policy issues may suddenly diverge, making for an intense debate or confrontation, while at the same time the opinions of individual citizens evolve gradually. Similarly, public policies may shift abruptly, although the direction of public opinion has changed gradually.

3. *Incremental delay.* In attempting to maintain or build public support, policymakers tend to choose policies that involve incremental changes of existing routines and practices. Incremental choice involves "the continual successive comparison of the adopted policy with all nearby alternatives."[28] Delay is a consequence of many factors:

[28]Isnard and Zeeman, "Some Models from Catastrophe Theory," p. 52. Incremental delay is normally labeled the "Delay Rule" (change policy in the direction that locally increases support). Other rules are "Maxwell's Rule"

incomplete information, the prevalence of intuitive (ordinary common sense) forms of analysis, political loyalties and commitments, institutional inertia, and historical precedent.

4. *Catastrophic policy change.* Incremental policymaking delays catastrophic change until the last possible moment, partly because smoothly evolving changes in the opinion of contending groups do not appear to require abrupt changes of direction. At a given point in time, policymakers are compelled to make sudden and discontinuous shifts in policy in order to retain popular support.

So far, the primary application of catastrophe methodology in public policy analysis has been to the area of public opinion.[29] A sudden and discontinuous shift in German energy policy, for example, has been explained by using catastrophe methodology to show how gradual changes in public opinion, together with incremental delay, produced a decision to suddenly terminate plans for the construction of a nuclear power plant in Kaiserstuhl, an agricultural area situated on the Rhine River in the state of Baden Wurttemberg.[30] Following public hearings, construction was begun on the plant, only to be followed by sit-ins and the forcible removal of demonstrators from the site. Throughout this period, public opinion gradually shifted in favor of the farming population in the region and its opposition to the plant. After numerous incremental policy changes and considerable cost to the government, the project was abruptly abandoned.[31]

Catastrophe methodology provides concepts and techniques for understanding discontinuous policy processes. However, it is a set of concepts and methods, rather than a theory. As such, it rests on the assumption that discontinuous processes observed in the past will simply repeat themselves in the future. Although catastrophe methodology attempts to provide sound theoretical reasons for the occurrence of future events (as contrasted with simple beliefs that patterns observed in the past will repeat themselves in the future), it is best viewed as a form of extrapolative forecasting. It is more a way to think about discontinuous policy futures than to make predictions derived from theory, for example, chaos theory.

THEORETICAL FORECASTING

Theoretical forecasting methods help analysts make *predictions* of future societal states on the basis of theoretical propositions and current and historical data. In contrast to extrapolative forecasting, which uses assumptions about historical recurrence to make forecasts, theoretical forecasting is based on assumptions about cause and effect contained in various theories. Whereas the logic of extrapolative forecasting is essentially inductive, the logic of theoretical forecasting is essentially deductive.

(change policy to where support is maximum) and the "Voting Rule" (change policy to where the *base,* rather than the maximum, support is greatest).

[29]See Isnard and Zeeman, "Some Models from Catastrophe Theory," pp. 45–60.

[30]See Rob Coppock, "Decision-Making When Public Opinion Matters," *Policy Sciences,* 8 (1977): 135–146.

[31]For a similar case by Jones, who uses the concept of "speculative augmentation," see Charles O. Jones, *Clean Air* (Pittsburgh, PA: University of Pittsburgh Press, 1975). Interestingly, Jones challenges the explanatory value of "disjointed incrementalism" (see Chapter 2), whereas Isnard and Zeeman see it as a major factor contributing to catastrophic change.

Deductive logic is a form of reasoning whereby certain general statements (axioms, laws, propositions) are used to justify the truth or falsity of other more specific statements, including predictions. Deductive reasoning is most frequently used in connection with arguments from cause that seek to establish that if one event (X) occurs, another event (Y) will follow. Although the distinguishing feature of theoretical forecasting is that predictions are deduced from theory, it should be emphasized that deduction and induction are interrelated. The persuasiveness of a deductive argument is considerably increased if theoretically deduced predictions are observed recurrently through empirical research. Similarly, an isolated empirical generalization ("The enemy chose to withdraw when threatened") is much more persuasive if backed by one or more assumptions contained in a theory ("The greater the costs of an alternative, the less likely that it will be chosen").[32]

In this section, we consider several procedures that assist analysts in making theoretical forecasts: theory mapping, causal modeling, regression analysis, point and interval estimation, and correlational analysis. Some of these procedures (e.g., theory mapping) are concerned with ways to identify and systematize theoretical assumptions, whereas others (e.g., regression) provide better estimates of future societal states predicted from theory. Before we begin, however, it is essential to recognize that none of these techniques actually makes predictions; only theory can make predictions.

Theory Mapping

Theory mapping is a technique that helps analysts identify and arrange key assumptions within a theory or causal argument.[33] Theory mapping can assist in uncovering four types of causal arguments: convergent, divergent, serial, and cyclic. *Convergent arguments* are those in which two or more assumptions about causation are used to support a conclusion or claim. *Divergent arguments* are those in which a single assumption supports more than one claim or conclusion. By contrast, in *serial arguments*, one conclusion or claim is used as an assumption to support a series of further conclusions or claims. Finally, *cyclic arguments* are serial arguments in which the last conclusion or claim in a series is connected with the first claim or conclusion in that series. The consequences of a cyclic argument may be positively or negatively self-reinforcing. These four types of causal arguments are illustrated in Figure 4.12.

A theory may contain a mixture of convergent, divergent, serial, and cyclic arguments. Several procedures may be used to uncover the overall structure of an argument or theory: (1) separate and number each assumption, which may be an axiom, law, or proposition; (2) *italicize* the words that indicate claims (*therefore, thus, hence*) or assumptions used to warrant claims (*since, because, for*); (3) when specific words (*therefore* and so on) have been omitted, but are clearly implied, supply the appropriate logical indicators in brackets; and (4) arrange numbered assumptions and claims in an arrow diagram that illustrates the structure of the causal argument or theory.

[32]The classic discussion of relations between theory and empirical data in the social sciences is Robert K. Merton, *Social Theory and Social Structure,* rev. ed. (Glencoe, IL: Free Press, 1957), pp. 95–99.

[33]Adapted from John O'Shaughnessy, *Inquiry and Decision* (New York: Harper & Row, 1973), pp. 58–61; and M. C. Beardsley, *Thinking Straight* (Englewood Cliffs, NJ: Prentice Hall, 1950).

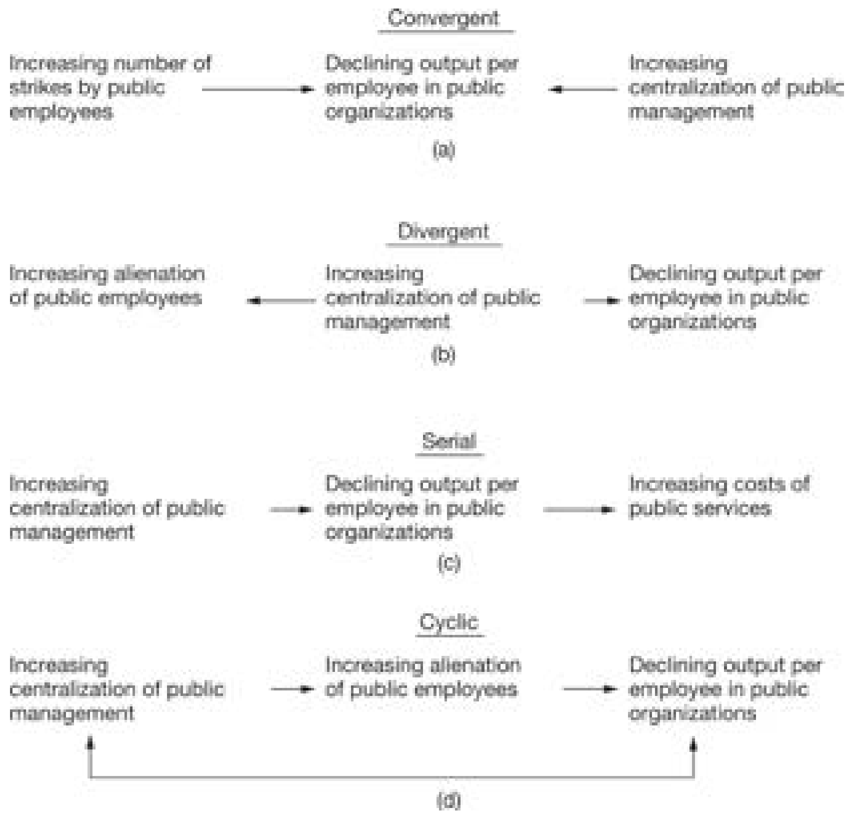


FIGURE 4.12
Four Types of Causal Arguments

Let us apply these procedures to an important theory about public policymaking. This theory, called "public choice," is concerned with the institutional conditions underlying efficient (and inefficient) government administration.[34] Public choice theorists have attempted to show on theoretical grounds how various institutional arrangements, particularly democratic forms of administration, will result in greater efficiency and public accountability. Public choice theory is also controversial, because it suggests that alleged problems of American government (inefficiency, inequity, unresponsiveness) are partly a consequence of the teachings of public administration, a discipline that for 50 and more years has emphasized centralization, hierarchy, and the consolidation of administrative powers.

In an effort to explain the inefficiency of governmental institutions, one public choice theorist, Vincent Ostrom, offers the following argument, which has been mapped by supplying logical indicators in brackets:

[34]See Vincent Ostrom, *The Intellectual Crisis in American Public Administration,* rev. ed. (University: University of Alabama Press, 1974).

> Gordon Tullock in *The Politics of Bureaucracy* (1965) analyzes the consequences which follow when [IT IS ASSUMED THAT] rational, self-interested individuals pursue maximizing strategies in very large public bureaucracies. Tullock's "economic man" is [THEREFORE] an ambitious public employee who seeks to advance his career opportunities for promotions within a bureaucracy. *Since* career advancement depends upon favorable recommendations by his superiors, a career-oriented public servant will act so as to please his superiors. [THEREFORE] Favorable information will be forwarded; unfavorable information will be repressed. [SINCE] Distortion of information will diminish control and create expectations which diverge from events generated by actions. Large-scale bureaucracies will *thus* become error-prone and cumbersome in adapting to rapidly changing conditions. [SINCE] Efforts to correct the malfunctioning of bureaucracy by tightening control will simply magnify errors. A decline in return to scale can [THEREFORE] be expected to result. [BECAUSE] The larger the organization becomes, the smaller the percent of its activities will relate to output and the larger the proportion of its efforts will be expended on management.[35]

In the passage just quoted, we used procedures (2) and (3) of theory mapping. We have also italicized words that indicate assumptions underlying claims and supplied missing logical operators in brackets. Observe that the argument begins with an assumption about human nature ("economic man"), an assumption so fundamental that it is called an *axiom*. Axioms are regarded as true and self-evident; that is, they are believed to require no proof. Note also that the overall structure of this complex argument is difficult to grasp simply by reading the passage.

The structure of the argument can be more fully described when we complete steps (1) and (4). Using procedure (1), we separate and number each claim and its warranting assumption, changing some of the words to improve the clarity of the original argument.

[SINCE]	1. Public employees working in very large public bureaucracies are rational, self-interested individuals who pursue maximizing strategies ("economic man").
[AND]	2. The desire for career advancement is a consequence of rational self-interest.
[THEREFORE]	3. Public employees strive to advance their career opportunities.
[SINCE]	4. The advancement of career opportunities is a consequence of favorable recommendations from superiors.
[AND]	5. Favorable recommendations from superiors are a consequence of receiving favorable information from subordinates.
[THEREFORE]	6. Subordinates striving to advance their careers will forward favorable information and suppress unfavorable information.
[SINCE]	7. The repression of unfavorable information creates errors by management, reducing their flexibility in adapting to rapidly changing conditions.
[AND]	8. The repression of unfavorable information diminishes managerial control.

[35]Ibid., p. 60.

[THEREFORE] 9. Managers compensate for their loss of control by attempting to tighten control. Compensatory attempts to tighten control further encourage the repression of unfavorable information (6) and the magnification of management errors (7). Further (8) loss of control produces attempts to tighten control.

[AND] 10. Compensatory tightening of control is a consequence of the size of public bureaucracies.

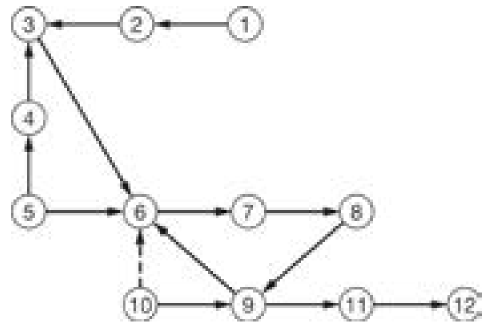
[SINCE] 11. Compensatory tightening of control requires the expenditure of a larger proportion of effort on management activities and a smaller proportion of output activities.

[THEREFORE] 12. The larger the public bureaucracy, the smaller the proportion of effort expanded on output activities in relation to size, that is, the less the return to scale.

By separating and numbering each claim and warranting assumption, we have begun to expose the logical structure of the argument. Note that there are different types of causal arguments in this theory of governmental inefficiency. The first part of the quotation contains a serial argument that emphasizes the influence of employee motivation on information error and managerial control. The second part of the quotation contains another serial argument, but one that stresses the importance of the size of public bureaucracies. In addition, there is one cyclic argument (error magnification) and several divergent and convergent arguments.

The structure of the overall argument cannot be satisfactorily exposed until we complete procedure (4) and draw an arrow diagram that depicts the causal structure of the argument (Figure 4.13). Observe, first, that there are two serial arguments: (1, 2, 3) and (5, 4, 3). The second of these (5, 4, 3) can stand by itself and does not require the first (1, 2, 3), which rests exclusively on deductions from the axiom of "economic man." Second, there is one divergent argument (5, 4, 6) that indicates that the same factor (5) has multiple consequences. There are also three convergent arguments (4, 2, 3), (3, 5, 6), and (8, 10, 9), which suggest in each case that there are two factors that explain the occurrence of the same event.

The arrow diagram also exposes two central features of this theoretical argument. It makes explicit an important potential relationship, denoted by the broken line between (10) and (6), which suggests that the size of public bureaucracies may independently affect the tendency to forward or suppress information. At the same time, the arrow diagram helps identify a cyclic argument (6, 7, 8, 9) that is crucial to this theory. This part of the theory argues, in effect, that we may expect a cumulative increase in the amount of management error as a result of the self-reinforcing relationship among information suppression (6), management error (7), loss of management control (8), compensatory tightening of control (9) further encouragement of information suppression (6), and so on. This cyclic argument, also known as a *positive feedback loop* (positive because the values of variables in the cycle continue to grow), suggests that any forecast based on public choice theory will predict a curvilinear pattern of growth in management error and government inefficiency.



1. Rational Self-Interest/Economic Man
2. Desire for Career Advancement
3. Advance Career Opportunities
4. Obtain Favorable Recommendations from Superiors
5. Obtain Favorable Information from Subordinates
6. Suppress Unfavorable Information
7. Commit Management Errors
8. Lose Managerial Control
9. Compensate by Tightening Control
10. Size of Public Bureaucracies
11. Expenditures on Management
12. Returns to Scale

FIGURE 4.13
Arrow Diagram Illustrating the Causal Structure of an Argument

If we had not used theory-mapping procedures, it is doubtful that we would have uncovered the convergent, divergent, serial, and cyclic arguments embedded in the structure of the theory. Particular causal assumptions—for example, the assumption that the size of public bureaucracies may combine with compensatory tightening of control to repress unfavorable information—are not clearly stated in the original passage. Because theory mapping makes the structure of claims and assumptions explicit, it provides opportunities to use methods of theoretical modeling.

Theoretical Modeling

Theoretical modeling refers to a broad range of methods for constructing simplified representations (models) of theories. Modeling is an essential part of theoretical forecasting, because analysts seldom if ever make theoretical forecasts directly from theory. Whereas analysts begin with theories, they develop models of these theories before they forecast future events. The process of testing a theory involves constructing and testing models of theories, not the theories themselves.

In the last chapter, we compared and contrasted models in terms of their aims (descriptive, normative); forms of expression (verbal, symbolic, procedural); and methodological functions (surrogates, perspectives). The majority of theoretical forecasting models are primarily descriptive, because they seek to predict rather than optimize some valued outcome. For the most part, these models are also expressed symbolically, that is, in the form of mathematical symbols and equations. Note that we have already used several symbolic models in connection with extrapolative forecasting, for example, the regression equation or model

[$Y_t = a + b\ (x)$]. Although it is not a causal model (because no explicit causal arguments are offered), it is nevertheless expressed symbolically.

In public policy analysis, a number of standard symbolic models assist in making theoretical forecasts: causal models, linear programming models, input-output models, econometric models, microeconomic models, and system dynamics models.[36] As it is beyond the scope of this book to detail each of these models, we confine our attention to causal models.

Causal Modeling

Causal models are simplified representations of theories that attempt to explain and predict the outcomes of public policies. The basic assumption of causal models is that correlation between two or more variables—for example, a correlation showing that increases in per capita income are associated with increases in welfare expenditures—are a reflection of underlying generative powers (causes) and their consequences (effects). The relation between cause and effect is expressed by laws and propositions contained within a theory and then modeled. Returning to the illustration from public choice theory, the statement "The proportion of total effort invested in management activities is determined by the size of public organizations" is a theoretical proposition. A model of that proposition might be $Y = a + b\ (X)$, where Y is the ratio of management to non-management personnel, a and b are constants, and X is the total number of employees in public organizations of different sizes.

The strength of causal models is that they force analysts to make causal assumptions explicit. The limitation of causal models lies in the tendency of analysts to confuse correlations uncovered through statistical analysis with causality.[37] Causal inferences always come from *outside* a model, that is, from laws, propositions, or assumptions within some theory. In the words of Sewall Wright, one of the early pioneers in this type of modeling, causal modeling procedures are "not intended to accomplish the impossible task of deducing causal relations from the values of the correlation coefficients."[38]

Causal modeling has been used to identify the economic, social, and political determinants of policies in issue areas ranging from transportation to health, education, and

[36]For a discussion of these and other standard models, see Martin Greenberger and others, *Models in the Policy Process* (New York: Russell Sage Foundation, 1976), Chapter 4; and Saul I. Gass and Roger L. Sisson, eds., *A Guide to Models in Governmental Planning and Operations* (Washington, DC: U.S. Environmental Protection Agency, 1974). Whereas these models are here treated as descriptive and symbolic (using the definitions of these terms provided in Chapter 3), some (e.g., linear programming) are often treated as normative. Similarly, others (e.g., system dynamics) are typically treated as procedural (or simulation) models. Distinctions among types of models are relative and not absolute.

[37]A recent systematic treatise on this issue is Judea Pearl, Madelyn Glymour, and Nicholas P. Jewell, *Causal Inference in Statistics: A Primer* (John Wiley & Sons, 2016). Among other things, they show that the language and notation of statistics provides no way to express causality. For this reason, they develop a new supplemental causal language.

[38]Sewall Wright, "The Method of Path Coefficients," *Annals of Mathematical Statistics* 5 (1934): 193, quoted in Fred N. Kerlinger and Elazar J. Pedhazur, *Multiple Regression in Behavioral Research* (New York: Holt, Rinehart and Winston, 1973), p. 305.

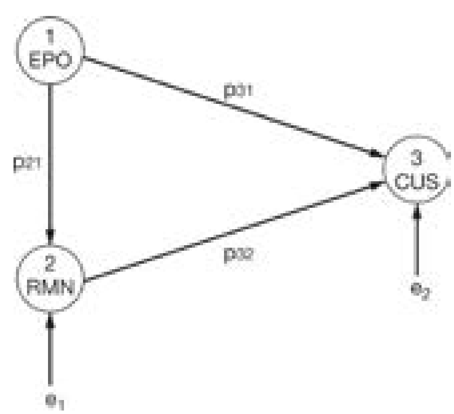
welfare.[39] One of the major claims of research based on causal modeling is that differences in political structures (e.g., single versus multiparty systems) do not directly affect such policy outputs as education and welfare expenditures. On the contrary, differences in levels of socioeconomic development (income, industrialization, urbanization) determine differences in political structures, which in turn affect expenditures for education and welfare. This conclusion is controversial because it appears to contradict the commonly shared assumption that the content of public policy is determined by structures and processes of politics, including elections, representative mechanisms, and party competition.

One of the main statistical procedures used in causal modeling is *path analysis*, a specialized approach to regression that uses multiple independent variables. Path analysis identifies those independent variables (e.g., income) that singly and in combination with other variables (e.g., political participation) determine changes in a dependent variable (e.g., welfare expenditures). An independent variable is the presumed cause of a dependent variable, which is presumed to be its effect. Estimates of cause and effect are called *path coefficients*, which express one-directional causal relationships among independent and dependent variables.

A standard way of depicting causal relationships is the path diagram. A path diagram looks very much like the arrow diagram used to map public choice theory, except that a path diagram contains estimates of the strength of the effects of independent variables on dependent variables. A path diagram has been used to model part of public choice theory in Figure 4.14. The advantage of path analysis and causal modeling is that they permit forecasts that are based on explicit theoretical assumptions about presumed causes (the number of employees in individual public organizations) and presumed effects (the ratio of managerial to non-managerial staff and costs in tax dollars per unit of public service). The limitation of these procedures, as already noted, is that they are not designed for the impossible task of inferring causation from estimates of relationships among variables. Although

FIGURE 4.14
Path Diagram Illustrating a Model of Public Choice Theory

Note: The symbol p designates a causal path and the subscripts specify the direction of causation: p_{31} means that variable 3 is caused by variable 1. Variables that have no antecedent cause are called exogenous variables (i.e., their cause is external to the system), whereas all others are called endogenous (i.e., their cause is internal to the system). The symbol e (sometimes designated as u) is an error term, defined as the unexplained (residual) variance in an endogenous variable that is left over after taking into account the effects of the endogenous variable that precedes it in the path diagram. Error terms should be uncorrelated with each other and with other variables.



[39]For a review, see Thomas R. Dye and Virginia H. Gray, "Symposium on Determinants of Public Policy: Cities, States, and Nations," *Policy Studies Journal* 7, 4 (Summer 1979): 279–301.

the absence of a relationship is sufficient to infer that causation is *not* present, only theory permits causal inferences and, hence, predictions.[40]

Regression Analysis

A useful technique for estimating relationships among variables in theoretical forecasting models is regression analysis. *Regression analysis*, which has already been considered in slightly modified form in our discussion of trend estimation, is a general statistical procedure that yields estimates of the pattern and magnitude of a relationship between a dependent variable and one or more independent variables. When regression analysis is performed with one independent variable, it is called *simple regression*; if there are two or more independent variables, it is called *multiple regression*. While many theoretical forecasting problems require multiple regression, we limit ourselves to simple regression in the remainder of this section.[41]

Regression analysis is particularly useful in theoretical modeling. It provides summary measures of the pattern of a relationship between an independent and dependent variable. These summary measures include a regression line, which permits us to estimate values of the dependent variable simply by knowing values of the independent variable, and an overall measure of the vertical distances of observed values from the regression line. A summary measure of these distances, as we shall see, permits us to calculate the amount of error contained in a forecast. Because regression analysis is based on the principle of *least squares,* it has a special advantage already noted in connection with linear trend estimation. It provides the one best "fit" between data and the regression line by ensuring that the squared distances between observed and estimated values represent a minimum or least value.[42]

A second advantage of regression analysis is that it compels the analyst to decide which of two (or more) variables is the cause of the other, that is, to specify the independent (cause) and dependent (effect) variable. To make decisions about cause and effect, however, analysts must have some theory as to why one variable should be regarded as the cause of another. Although regression analysis is particularly well suited to problems of predicting effects from causes, the best that regression analysis can do (and it does this well) is to provide estimates of relations predicted by a theory. Yet it is the theory and its simplified representation (theoretical model), and not regression analysis, that do the predicting. Regression analysis can only provide estimates of relations between variables that, because

[40]The exception to this statement is *descriptive causality* established by means of experimentation, which requires that an intervention, the presumed cause, temporally precede an outcome, the presumed effect. Other means of establishing causality are the elimination of rival causes through randomization and the use of a control group, which permits counterfactual claims. The authoritative source on this question is William R. Shadish, Thomas D. Cook, and Donald T. Campbell, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference* (Belmont, CA: Wadsworth, Cengage Learning, 2002).

[41]For a thorough and readable classical treatment of multiple regression and related techniques, see Kerlinger and Pedhazur, *Multiple Regression in Behavioral Research.* Recent texts include J. H. Stock and M. W. Watson, *Introduction to Econometrics* (Boston: Addison Wesley, 2003).

[42]If this point is not clear, you should return to the section on extrapolative forecasting and review Figure 4.8.

of some theory, have been stated in the form of predictions.[43] For this reason, analysts should employ theory-mapping procedures before using regression analysis.

To illustrate the application of regression analysis to problems of theoretical forecasting, let us suppose that municipal policymakers wish to determine the future maintenance costs of police patrol vehicles under two alternative policies. One policy involves regular police patrols for purposes of traffic and crime control. In 1980, the total maintenance costs for the ten patrol cars were $18,250, or $1,825 per vehicle. The total mileage for the ten vehicles was 535,000 miles, that is, 53,500 miles per vehicle. A new policy now being considered is one that would involve "high-impact" police patrols as a way to create greater police visibility, respond more rapidly to citizens' calls for assistance, and, ultimately, deter crime by increasing the probability that would-be offenders are apprehended.[44]

Local policymakers are interested in any means that will reduce crime. Yet there are increasing gaps between revenues and expenditures, and several citizens' groups have pressed for a cutback in municipal employees. Policymakers need some way to forecast, on the basis of their own mileage and maintenance records, how much it will cost if several of the ten vehicles are driven an additional 15,000 miles per year.

This forecasting problem may be effectively dealt with by using regression analysis. The relation between cause and effect is reasonably clear, and the primary determinant of maintenance costs is vehicle usage as measured by miles driven.[45] A municipal policy analyst might therefore plot the values of the independent (X) and dependent (Y) variables on a *scatter diagram* (*scatterplot*), which will show the pattern of the relation (linear–nonlinear), the direction of the relation (positive–negative), and the strength of the relation (strong–moderate–weak) between annual mileage per vehicle and annual maintenance costs per vehicle (Figure 4.15). We assume that the pattern, direction, and strength of the relationship is linear, positive, and strong, as shown in Figure 4.15(a).

Linear regression analysis assumes that variables are related in a linear pattern. While linear regression analysis may also be used with negative linear relationships (Figure 4.15[b]), it will produce serious errors if applied to curvilinear patterns such as those illustrated in Figures 4.15(c) and (d). In such cases, as we saw earlier in the discussion of curve fitting, we must either use nonlinear regression analysis or transform values of the variables (e.g., by taking their logarithms) prior to applying conventional linear regression techniques. Regression analysis will yield less reliable estimates when data are widely

[43]Recall the distinction between surrogate and perspective models in Chapter 3 and the example of the annual rainfall rate and reservoir depth. This and other examples show vividly that regression analysis, apart from its superiority in making estimates, cannot answer questions about which variable predicts another.

[44]In the 1970s, under the sponsorship of the Law Enforcement Assistance Administration, high-impact patrolling was used, with mixed success, in a number of large municipalities. See Elinor Chelimsky, "The Need for Better Data to Support Crime Control Policy," *Evaluation Quarterly* 1, 3 (1977): 439–474.

[45]The question of causation is never certain, as illustrated by this example. Under certain conditions, maintenance costs may affect miles driven, for example, if cost-conscious managers or policemen limit vehicle use in response to knowledge of high expenses. Similarly, annual mileage for certain vehicles may mean larger patrol areas, which in turn may be situated where road conditions are better (or worse) for cars.

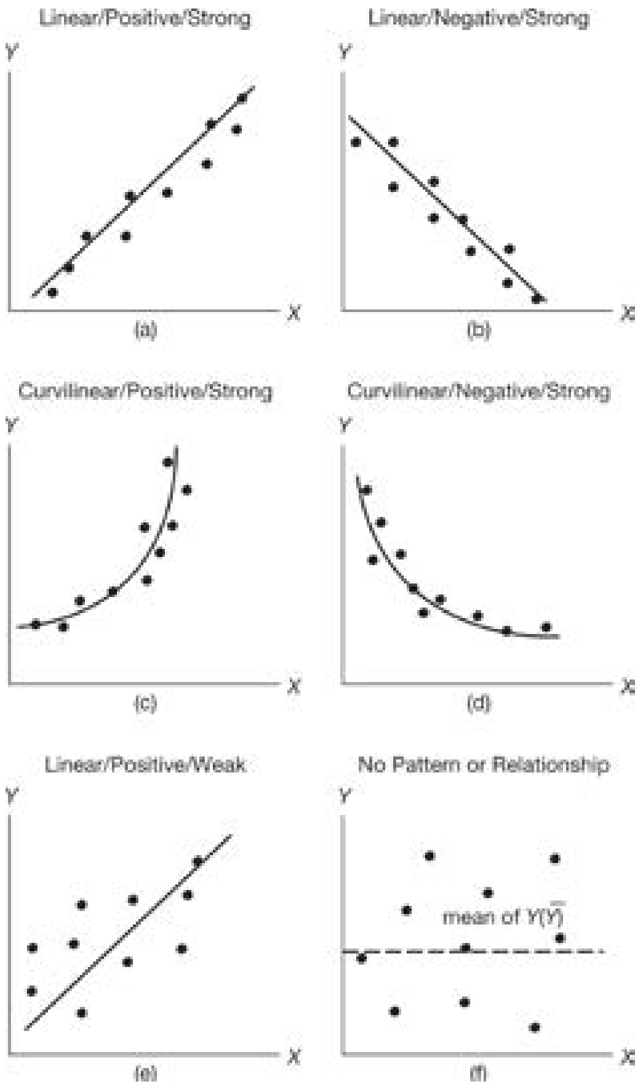


FIGURE 4.15
Scatterplot Illustrating Different Patterns and Relations between Hypothetical Annual Maintenance Costs and Annual Mileage per Vehicle

scattered, as in Figure 4.15(e). If data indicate no pattern or relation (Figure 4.15[f]), the best estimate of changes in Y due to X is the average (mean) of values of Y.

Recall from our discussion of trend estimation that the straight line describing the relationship between X and Y variables is called a *regression line*. The equation used to fit the regression line to observed data in the scatter diagram is identical to that used to estimate linear trend, with several minor differences. The symbol Y_t (the subscript t refers to a trend value) is replaced with Y_c (the subscript c refers to a computed, or estimated, value). These different subscripts remind us that here, regression is applied to two substantive variables, whereas in linear trend estimation one of the variables is time. The formula for the regression equation is

$$Y_c = a + b(X)$$

where

a = the value of Y_c when $X = 0$, called the *Y intercept* because it shows where the computed regression line intercepts the Y-axis
b = the value of changes in Y_c due to a change of one unit in X, called the *slope* of the regression line because it indicates the steepness of the straight line
X = a given value of the independent variable

A second minor difference between the regression and trend equations is the computation of values of a and b. In regression analysis, we do not work with original values of Y and coded time values of X but with *mean deviations.* A mean deviation is simply the difference between a given value of X or Y and the average (mean) of all values of X or Y. The formulas necessary to calculate the mean and mean deviations for X and Y are:

$$\text{mean of } Y = \bar{Y} = \frac{\Sigma Y}{n}$$
$$\text{mean deviation of } Y = y = Y - \bar{Y}$$

The values of a and b in the regression equation are computed with the following formulas:

$$b = \frac{\Sigma(xy)}{\Sigma(x^2)}$$
$$a = \bar{Y} - b(\bar{X})$$

In the forecasting problem facing the municipal policy analyst, all data needed to solve the regression equation are provided in Table 4.6. Column (1) lists the ten vehicles in the municipal fleet. Annual mileage (X) and maintenance costs (Y) are expressed in thousands of miles and dollars [columns (2) and (3)]. At the bottom of columns (1) and (2), we have calculated the sums of X (X) and Y ($\Sigma\ Y$) and their respective means ($\bar{X} = \Sigma X/n = 53.6$ and $\bar{Y} = \Sigma Y/n = 1.785$). The deviations of each X and Y value from its respective mean are in columns (4) and (5). These mean deviations $(x = \bar{X} - X)$ and $(y = Y - \bar{Y})$ show the spread of each X and Y value about its own mean. Note that the values of the mean deviation sum to zero [bottom of columns (4) and (5)], because the values above the mean cancel those below the mean.

The mean deviations (x and y) exhibit the spread of each value around the X and Y means. Nevertheless, they cannot be used as a summary measure of this spread, because they sum to zero. For this reason, mean deviations are squared [columns (7) and (8)], thus eliminating negative signs. They are then summed. These sums, called the *sums of squares* (which is a shortened form of "sums of the squares of mean deviations"), provide an overall summary of the spread of X and Y values around their means. Finally, in column (6) we have multiplied the mean deviations (x and y), obtaining *deviation cross-products* (xy).

TABLE 4.6

Worksheet for Estimating Future Maintenance Costs from Annual Mileage per Vehicle

Vehicle Number	Mileage per Vehicle (Thousands of Miles)	Maintenance Costs per Vehicle (Thousands of Dollars)	X Mean Deviation $(X-\bar{X})$	Y Mean Deviation $(Y-\bar{Y})$	Deviation Cross-Product	X Mean Deviation Squared	Y Mean Deviation Squared	Estimated Values of
	(X)	(Y)	(x)	(y)	(xy)	(x^2)	(y^2)	(Y_c)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1	24	0.45	−29.5	−1.335	39.38	870.25	1.7820	0.838
2	23	1.25	−30.5	−0.585	17.84	930.25	0.3420	0.806
3	30	0.85	−23.5	−0.985	23.15	552.25	0.9700	1.030
4	40	1.15	−13.5	−0.685	9.25	182.25	0.4690	1.400
5	48	1.95	−5.5	0.115	−0.633	30.25	0.0130	1.606
6	65	1.85	11.5	0.015	0.173	132.25	0.0002	2.150
7	65	2.35	11.5	0.515	5.923	132.25	0.2650	2.150
8	60	2.65	6.5	0.815	5.298	42.25	0.6640	1.990
9	90	2.55	36.5	0.715	26.098	1332.25	0.5110	2.950
10	91	3.25	37.5	1.415	53.063	1406.25	2.0020	2.982
$n = 10$	$\Sigma X = 536$ $\bar{X} = 53.6$	$\Sigma Y = 17.85$ $\bar{Y} = 1.785$	0.0	0.0	$\Sigma(xy) = 179.54$	$\Sigma(x^2) = 5610.5$	$\Sigma(y^2) = 7.02$	$\Sigma Y_c = 17.85$

$$b = \frac{\Sigma(xy)}{\Sigma(x^2)} = \frac{179.54}{5610.5} = 0.032$$

$$a = \bar{Y} - b(\bar{X}) = 1.785 - 0.032(53.6) = 0.0698 \text{ } or \text{ } 0.07 \text{ (rounded)}$$

$$Y_C = a + b(X) = 0.07 + 0.032(X)$$

$$Y_{150} = 0.07 + 0.032(150) = 4.87 = \$4,870$$

These cross-products are an expression of the pattern and strength of the relationship between X and Y, but one that takes into account the spread of X and Y scores.

To estimate future maintenance costs under the high-impact police patrol program, we first solve the regression equation

$$Y_c = a + b(X)$$

by substituting calculations from the worksheet into formulas for a and b. Hence,

$$b = \frac{\Sigma(xy)}{\Sigma(x^2)} = \frac{179.54}{5610.5} = 0.032$$

$$a = \bar{Y} - b(\bar{X}) = 1.785 - 0.032(53.6) = 0.07$$

$$Y_c = 0.07 + 0.032(X)$$

This regression equation means that the regression line intercepts the Y-axis at 0.07 thousand dollars (i.e., \$70) and that maintenance costs increase by 0.032 thousand dollars (i.e., \$32) for every 1,000 miles driven. On the basis of this information, precise estimates of maintenance costs can be made on the basis of knowledge of annual miles per vehicle. Computed values (Y_c) for each original value of X [column (9)] may be plotted on a graph to create the regression line. Finally, our forecast indicates that the city will incur an additional \$4,870 in expenses by adopting high-impact patrolling, because an additional 150,000 miles will result in $Y_{150} = 0.07 + 0.032(150) = 4.87$ thousand dollars.

Point and Interval Estimation

Regression analysis enables us to fit a regression line to data in such a way that the squared distances between data points and the regression line are a minimum or least value (Figure 4.8). This means that regression analysis not only uses information about variations in costs and mileage; it also permits us to make a sound estimate of the central tendency in a relationship and compare this estimate with observed values (the dots) in a scatter diagram. Because the distances between the regression line and individual data points are minimized, the regression estimate contains less error than other kinds of estimates, for example, those calculated on the basis of average costs per mile driven. Regression analysis also allows us to calculate the probable error in our estimate. This may be done in several ways (Table 4.7). One way is simply to subtract each value of Y_c [column (3)] from its corresponding observed value [column (2)]. This tells us how much distance there is between each pair of estimated and observed values. The error of a regression estimate, expressed by the distance between values of Y and Y_c, increases in direct proportion to the amount of scatter. The greater the scatter, the greater the error, and the less accurate the estimate.

Another way to calculate error is to sum and average these distances. Because one of the properties of linear regression is that these distances equal zero when they are added up [i.e., $\Sigma(Y - Y_c) = 0$], we must square them before we divide by the number of cases to find the average. Column (5) of Table 4.7 shows that the average squared error (also called the *variance of an estimate*) is 1.304. This expression of error, although useful for certain calculations to be described shortly, is difficult to interpret. Fortunately, the square root of

TABLE 4.7

Calculation of Standard Error of Estimated Maintenance Costs

Vehicle Number (1)	Observed Maintenance Costs (Y) (2)	Estimated Maintenance Costs (Y_c) (3)	Observed Minus Estimated Costs ($Y - Y_c$) (4)	Observed Minus Estimated Costs Squared ($(Y - Y_c)^2$) (5)
1	0.45	0.838	−0.388	0.151
2	1.25	0.806	0.394	0.155
3	0.85	1.030	−0.230	0.053
4	1.15	1.400	−0.300	0.090
5	1.95	1.606	0.294	0.086
6	1.85	2.150	−0.350	0.123
7	2.35	2.150	0.150	0.023
8	2.65	1.990	0.610	0.372
9	2.55	2.950	−0.450	0.203
10	3.25	2.982	0.218	0.048
$n = 10$	$\Sigma Y = 17.85$	$\Sigma Y_c = 17.85$	$\Sigma (Y - Y_c) = 0$	1.304

$$S_{y \cdot x} = \sqrt{\frac{\Sigma(Y - Y_c)^2}{n}} = \sqrt{\frac{\Sigma(Y - Y_c)^2}{10}} = \sqrt{\frac{1.304}{10}} = 0.36111$$

$$Y_i = Y_c \pm Z\left(S_{y \cdot x}\right)$$

$$\begin{aligned} Y_{i(150)} &= 4.87 \pm 2(0.3611) = 4.87 \pm 0.72222 \\ &= 4.14778 \text{ to } 5.59222 \\ &= \$4{,}147.78 \text{ to } \$5{,}592.22 \end{aligned}$$

this value is a good approximation of the error that will occur about two-thirds of the time in a regression estimate.[46] The square root of the average squared error, called the *standard error of estimate,* is calculated with the following formula:[47]

$$S_{y \cdot x} = \sqrt{\frac{\Sigma(Y - Y_c)^2}{n}}$$

[46]For readers who recall the normal (bell-shaped) curve, the standard error is the standard deviation of the average squared error. One standard error is one standard deviation unit to the left or right of the mean of a normal distribution and therefore includes the middle 68.3 percent of all values in the distribution. In the example given, we have assumed for purposes of illustration that data are normally distributed, as in a bell-shaped curve.

[47]Note that we divide by n, because the ten vehicles constitute the entire population of vehicles under analysis. If we were working with a sample, we would divide by $n - 1$. This provides an unbiased estimate of the variance in the sample of the population. Most packaged computer programs use $n - 1$, although we may be working with a population and not a sample.

where

$(Y - Y_c)^2$ = the squared difference between observed and estimated values of the dependent variable
n = the number of cases

In the previous example, the standard error of estimate is

$$S_{y \cdot x} = \sqrt{\frac{\Sigma\left(Y - Y_c\right)^2}{n}} = \sqrt{\frac{1.304}{10}}$$
$$= \sqrt{.1304} = 0.3611 \text{ thousand}$$
$$= \$361.11 \text{ in annual maintenance costs}$$

This means that any actual value of maintenance costs is likely to be \$361.11 above or below any estimated value about two-thirds of the time. The figure of \$361.11 is one standard error (1 × \$361.11), two standard errors are \$722.22 (2 × \$361.11), and three standard errors are \$1,083.33 (3 × \$361.11). The standard error gives us a *probability interpretation* of these values, because one standard error unit will occur about two-thirds (actually 68.3 percent) of the time, two standard error units will occur about 95 percent (actually 95.4 percent) of the time, and three standard error units will occur about 99 percent (actually 99.7 percent) of the time.

The standard error of estimate allows us to make estimates that incorporate error. Rather than make simple *point estimates*—that is, estimates that produce a single value of Y_c—we can make *interval estimates* of Y_c expressed in terms of one or more standard units of error. If we want to be 95 percent confident that our estimate is accurate (this is usually regarded as a minimum standard of accuracy), we will want to express our estimate in terms of two intervals above and below the original point estimate. We can use the following formula to find out how much error we might expect in our point estimate of \$4,870 [$Y_{150} = 0.07 + 0.032(150) = 4.87$ thousand dollars] 95 percent of the time. Thus:

$$Y_i = Y_c + Z\left(S_{y \cdot x}\right)$$
$$= 4.87 \pm 2\left(0.3611\right)$$
$$= 4.87 \pm 0.72222 = 4.14778 \text{ to } 5.59222$$
$$= \$4{,}147.78 \text{ to } \$5{,}592.22$$

where

Y_c = the point estimate of Y
Z = a standard error unit, which takes the value of 2 (95% confidence) in this case
$S_{y \cdot x}$ = the value (0.3611) of one standard error unit
Y_i = an interval estimate of Y (±0.72222)

Correlational Analysis

Regression analysis has an additional feature of importance to theoretical forecasting: It permits us to use *correlational analysis* to interpret relationships. Recall that different scatter diagrams not only exhibit the pattern of a relationship but also its direction and strength

(Figure 4.15). Yet it is desirable that we have measures of the direction and strength of these relationships, not just the visual images contained in scatter diagrams. Two measures that yield such information can be calculated from the worksheet already prepared for estimating future maintenance costs (Table 4.7). The first of these measures is the *coefficient of determination* (r^2), which is a summary measure or index of the amount of variance in the dependent variable explained by the independent variable. The second is the *coefficient of correlation* (r), which is the square root of the coefficient of determination. The coefficient of correlation, which varies between −1.0 and +1.0, tells us whether the direction of the relationship is positive or negative and how strong it is. If r is 0, there is no relationship, whereas a value of ± 1.0 (i.e., positive or negative) indicates a maximal relationship (sometimes this is called "perfect" correlation).

The coefficient of determination (r^2), unlike the coefficient of correlation (r), takes a positive sign only and varies between 0.0 and 1.0. The formulas for both coefficients are applied here to our data on maintenance costs and mileage (see Table 4.7):

$$r^2 = \frac{b(\Sigma xy)}{\Sigma\left(y^2\right)} = \frac{0.032(179.54)}{7.02} = 0.818 \text{ or } 0.82$$

$$r = \sqrt{r^2} = \sqrt{0.818} = 0.90$$

The analysis carried out by hand in Tables 4.6 and 4.7 has been done with SPSS (Exhibit 4.2). SPSS produces identical results, but with greater efficiency and accuracy. When we inspect the SPSS output, we see that the correlation coefficient (r), designated in the output as R (rather than r), is 0.905. The coefficient of determination (r^2), designated as R SQUARE, is 0.818. All coefficients are in agreement with our hand calculations. The output also provides the value of the intercept (CONSTANT = 0.07), which conforms to the output in Table 4.7, as does the slope of the regression line (B = 0.032) for the variable annual miles per vehicle. Finally, the ANOVA (*analysis of variance*) table provides a breakdown of that part of the total sum of squares due to the effects of regressing the dependent variable on the independent variable (regression sums of squares = 5.746). The output also provides that part of the sum of squares due to error (residual sums of squares = 1.275). If we divide the regression plus residual sum of squares into the regression sum of squares, we obtain the coefficient of determination (R). Thus,

$$\frac{5.746}{5.746 + 1.275} = 0.818 = 0.82$$

These coefficients tell us that 82 percent of the variance in annual maintenance costs is explained by annual mileage (r^2 = 0.82) and that the direction and strength of the relationship between these two variables is positive and strong (r = 0.90). Regression analysis, when supplemented by interval estimation and coefficients of determination and correlation, provides more information of direct relevance to policy than other forms of estimation. For example, simple estimates of average costs per mile, apart from their relative inaccuracy, do not provide summary measures of the direction and strength of relationships. Nor do they incorporate forecasting errors in a systematic way. In this and many other cases, regression analysis can help policymakers deal with the uncertainties that accompany efforts to predict the future.

EXHIBIT 4.2

SPSS Output for Table 4.7

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	Annual Miles Per Vehicle (000s)[a]		Enter

[a]All requested variables entered
[b]Dependent Variable: Annual Maintenance Costs Per Vehicle ($000)

Model Summary

Model	*R*	*R* Square	Adjusted *R* Square	Std. Error of the Estimate
1	.905[a]	.818	.796	.39917721

[a]Predictors: (Constant), Annual Miles Per Vehicle (000s)

ANOVA[b]

Model		Sum of Squares	*df*	Mean Square	*F*	Sig.
1	Regression	5.746	1	5.746	36.058	.000[a]
	Residual	1.275	8	.159		
	Total	7.020	9			

[a]Predictors: (Constant), Annual Miles Per Vehicle (000s)
[b]Dependent Variable: Annual Maintenance Costs Per Vehicle ($000)

Coefficients[a]

		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	*t*	Sig.
1	(Constant)Annual Miles Per	6.973E-02	.312		.223	.829
	Vehicle (000s)	3.200E-02	.005	.905	6.005	.000

[a]Dependent Variable: Annual Maintenance Costs Per Vehicle ($000)

JUDGMENTAL FORECASTING

In extrapolative and theoretical forecasting, empirical data and/or theories play a central role. In contrast, judgmental forecasting attempts to elicit and synthesize informed judgments from experts and other “knowledgeables.” The logic of judgmental forecasting is partly abductive, because analysts begin with a claim about a valued state of affairs (e.g., a normative future such as world peace) and work backwards to the data or assumptions necessary to support the claim. Nevertheless, inductive, deductive, and abductive reasoning are not mutually exclusive in practice. For this reason, judgmental forecasting is often supplemented by extrapolative and theoretical forecasting procedures.[48]

In this section, we review three judgmental forecasting methods based on expertise: Delphi technique, cross-impact analysis, and feasibility forecasting. These and other methods, which have been widely used in government, industry, and medicine, are particularly well suited to the kinds of problems we described in Chapter 3 as messy, ill structured, or wicked. Because one of the characteristics of ill-structured problems is that policy alternatives and their consequences are unknown, it follows that there are no relevant theories and/or empirical data to make forecasts. Under these conditions, judgmental forecasting techniques are useful and even necessary.

The Delphi Technique

The *Delphi technique* is a judgmental forecasting method that develops informed beliefs about the future. The Delphi technique—named after Apollo's shrine at Delphi, where Greek oracles sought to foresee the future—was developed in 1948 by researchers at the RAND Corporation. Since then it has been used in thousands of forecasting projects in the public, private, and nonprofit sectors. Originally, Delphi technique was applied to problems of military strategy, but its application gradually shifted to forecasts in other contexts: education, technology, marketing, transportation, mass media, medicine, information processing, research and development, space exploration, housing, budgeting, and the quality of life.[49] Whereas the technique originally emphasized the use of experts to verify forecasts based on empirical data, Delphi began to be applied to problems of values forecasting in the 1960s.[50] The Delphi technique has been used in countries ranging from the United States, Canada, and the United Kingdom to Japan and Russia.

Early applications of Delphi were motivated by a concern with the ineffectiveness of committees, expert panels, and other group processes in making sound judgments. The technique was designed to avoid several sources of distorted communication found in groups: domination of the group by one or several persons, pressures to conform to peer group opinion, personality differences and interpersonal conflict, and the difficulty of publicly opposing persons in positions of authority. To avoid these problems, early applications of Delphi were based on five basic principles:

1. S*tatistical group anonymity*. Experts or knowledgeables respond as individuals whose anonymity is preserved by membership in a statistical group.

[48]Ascher shows that large-scale econometric models also depend on judgment. See Ascher, *Forecasting;* Ascher, "The Forecasting Potential of Complex Models"; and McNown, "On the Use of Econometric Models."
[49]Delphi and other judgmental forecasts are reviewed by Harold Sackman, *Delphi Critique* (Lexington, MA: D. C. Heath and Company, 1975); Juri Pill, "The Delphi Method: Substance, Contexts, a Critique and an Annotated Bibliography," *Socio-Economic Planning Sciences* 5 (1971): 57–71; Harold A. Linstone, ed., "Forecasting: A New Agenda," in *Technological Forecasting and Social Change: An International Journal* 36, 1–2 (August 1989), pp. 1–234; and Nicholas Rescher, *Predicting the Future: An Introduction to the Theory of Forecasting* (Albany: State University of New York Press, 1998).
[50]See, for example, Nicholas Rescher, *Delphi and Values* (Santa Monica, CA: RAND Corporation, 1969).

2. *Iteration.* Individual judgments are aggregated and communicated back to participating experts in a series of rounds, thus permitting social learning and the modification of prior judgments.
3. *Controlled feedback.* The communication of aggregated judgments occurs in the form of summary measures of responses to questions, for example, frequency distributions of responses.
4. *Statistical group response.* Summaries of individual responses are presented in the form of measures of central tendency (usually the median) and dispersion (the semi-interquartile range), and frequency distributions (histograms and frequency polygons); and
5. *Expert consensus.* The central aim is to create conditions under which a consensus among experts is likely to emerge as the final product.

These principles represent a characterization of c*onventional Delphi,* which dominated the field well into the late 1960s. It should be contrasted with *policy Delphi,* a response to limitations of conventional Delphi and an attempt to create new procedures that match the complexities of policy problems. In the words of one of its chief architects,

> Delphi as it originally was introduced and practiced tended to deal with technical topics and seek a consensus among homogeneous groups of experts. The Policy Delphi, on the other hand, seeks to generate the strongest possible opposing views on the potential resolution of a major policy issue . . . a policy issue is one for which there are no experts, only informed advocates and referees.[51]

Policy Delphi is based on two of the same principles as conventional Delphi (iteration and controlled feedback), but it also introduces several new ones:

1. *Selective anonymity.* Participants in a policy Delphi remain anonymous only during the first or second round. After contending arguments about policies have surfaced, participants are asked to debate their views publicly.
2. *Informed multiple advocacy.* The process for selecting participants is based on criteria of commitment and knowledge, rather than "expertise" per se. In forming a Delphi group, investigators attempt to select a representative group of informed advocates.
3. *Polarized statistical response.* In summarizing individual judgments, measures that purposefully accentuate disagreement and conflict are used. Although conventional measures may also be used (median and semi-interquartile range), policy Delphi supplements these with various measures of variation and difference among individuals.
4. *Structured conflict.* Starting from the assumption that conflict is a normal feature of policy issues, every attempt is made to use disagreement for creatively exploring alternatives and their consequences. In addition, efforts are made to surface and

[51]Murray Turoff, "The Design of a Policy Delphi," *Technological Forecasting and Social Change* 2, 2 (1970): 149–171. See also Harold A. Linstone and Murray Turoff, eds., *The Delphi Method: Techniques and Applications* (New York: Addison-Wesley, 1975).

make explicit the assumptions and arguments that underlie contending positions. The outcomes of a policy Delphi are nevertheless completely open, which means that consensus as well as a continuation of conflict might be results of the process.

5. *Computer conferencing.* When possible, computer consoles are used to structure a continuous process of anonymous interaction among physically separated individuals. Computer conferencing may eliminate the need for a series of Delphi rounds.

A policy Delphi may be conducted in a number of ways, depending on the context and the skill of the persons using the technique. Because policy Delphi is a major research undertaking, it involves a range of technical tasks, including sampling, questionnaire design, reliability and validity, and data analysis and interpretation. Although these questions are beyond the scope of this chapter,[52] it is important to obtain an overall understanding of the process of conducting a policy Delphi, which can best be visualized as a series of interrelated steps:[53]

Step 1: *Issue specification.* The forecaster must decide which specific issues should be addressed. For example, if the area of concern is national drug abuse policy, one of the issues might be "The personal use of marijuana should be legalized." An important problem of this step is deciding what proportion of issues should be generated by participants, and what proportion should be generated by the forecaster. If the forecaster is well informed about the issue area, a list of issues may be developed prior to the first round of the Delphi. These issues may be included in the first questionnaire, although respondents should be free to add or delete issues.

Step 2: *Selection of advocates.* Key stakeholders should be selected. To select a group of stakeholders who represent conflicting positions it is desirable to use explicit sampling procedures such as "snowball" sampling. The forecaster begins by identifying one advocate, usually someone who is known to be influential in the issue area, and asking that person to name two others who agree and disagree with his or her own position. These two persons are asked to do the same thing, which results in two more persons who agree and disagree, and so on (hence the term *snowball sample*). The issue positions of stakeholders should be as different as possible in terms of their relative influence, formal authority, and group affiliation. The size of the sample might range from ten to thirty persons, although this depends on the nature of the issue. The more complex the issue, and hence the more heterogeneous the participants, the larger the sample must be to be representative of the positions of stakeholders.

---

[52]There are few shortcuts to developing methodologically sound questionnaires for use in policy Delphis. On questionnaire construction, the best short introduction is Earl R. Babbie, *Survey Research Methods* (Belmont, CA: Wadsworth, 1973). On reliability and validity, see Fred N. Kerlinger and Howard B. Lee, *Foundations of Behavioral Research,* 4th ed. (New York: Cengage Learning, 2000), pp. 599–686. A useful general-purpose handbook on these questions is Delbert C. Miller and Neil. J. Salkind, *Handbook of Research Design and Social Measurement,* 6th ed. (Newbury Park, CA: Sage Publications, 2002).

[53]See Turoff, "The Design of a Policy Delphi," pp. 88–94.

Step 3: *Questionnaire design.* Because a policy Delphi takes place in a series of two or more rounds, forecasters must decide which items will go in questionnaires to be used in the first and subsequent rounds. However, the second-round questionnaire can be developed only after the results of the first round are analyzed; the third-round questionnaire must await results of the second round; and so on. For this reason, only the first-round questionnaire can be drafted in advance. The first-round questionnaires may be relatively unstructured, with open-ended items, or relatively structured, provided the forecaster has a good idea of the major issues.

First-round questionnaires may include several types of questions: (1) *Forecasting items* that ask for subjective estimates of the probability of occurrence of particular events; (2) *Issue items* requesting respondents to rank issues in terms of their importance; (3) G*oal items* that solicit judgments about the desirability of pursuing certain goals; and (4) *Feasibility items* asking respondents for judgments about the feasibility of options for achieving goals.

Several types of scales are available to measure responses to different types of items. For example, a likelihood scale is appropriate for forecast items, an importance scale for issue items, desirability scales with goal items, and some combination of these scales with options items. Different types of items and scales are presented in the sample Delphi questionnaire displayed in Table 4.8.

The scales in Table 4.8 do not permit neutral answers. This is designed to sharpen disagreement and avoid response set (responding down the middle), which are important aims of the policy Delphi. In addition, it is important to pretest the questionnaire by calculating the internal consistency reliability of individual scales and the set of scales as a whole. For this purpose, software programs for *Cronbach's alpha*, a measure of internal consistency reliability, are available in SPSS, Stata, and R. Reliability as well as validity may be assessed by using a multitrait, multimethod matrix.[54]

Step 4: *Analysis of first-round results.* When the questionnaires are returned after the first round, forecasters attempt to determine the initial positions on forecasts, issues, goals, and options. Typically, some items believed to be desirable or important are also believed to be unfeasible, and vice versa. Because there will be conflicting assessments among the various experts, it is important to use summary measures that not only express the central tendency of the responses but also their dispersion and polarization. These summary measures not only eliminate items that are uniformly important, undesirable, unfeasible and/or uncertain but also serve in the second-round questionnaire as a means to communicate to participants the results of the first round.

---

[54]Early critics of Delphi questionnaires rightly called for reliability and validity testing, for example, Harold Sackman, *Delphi Critique: Expert Opinion, Forecasting, and Group Process* (Lexington: Lexington Books and D. C. Heath, 1975). Kerlinger and Lee provide a systematic but economical presentation of reliability and validity testing. Fred N. Kerlinger and Howard B. Lee. *Foundations of Behavioral Research,* 4th ed. (Cengage Learning, 2000). In addition, there is no substitute for using sources on competent questionnaire design, one of the best of which is Don A. Dillman, Jolene D. Smyth, and Leah Melani Christian. *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method* (New York: John Wiley & Sons, 2014).

TABLE 4.8

Types of Items and Scales Used in a Policy Delphi Questionnaire

Type of Item	Item	Probability Scale				
Forecasting	According to a 2015 survey of marijuana use, some 35 percent of 12th grade students used marijuana in the last 12 months, up from slightly more than 20 percent in 1991. What is the likelihood that marijuana use among 12th graders will exceed 40 percent by 2025? (please circle)	.0 .10 .20 .30 .40 .50 .60 .70 .80 .90 1.0				
Issue	The personal use of marijuana *should be* legalized.	Strongly Agree 1 [ ]	Agree 2 [ ]	Disagree 3 [ ]	Strongly Disagree 4 [ ]	No Judgment 0 [ ]
Goal	One goal of national policy might be to increase public awareness of the difference between responsible drug use and drug abuse. How *desirable* is this objective?	Very Desirable 1 [ ]	Desirable 2 [ ]	Undesirable 3 [ ]	Very Undesirable 4 [ ]	No Judgment 0 [ ]
Options	It has been suggested that the legalization of marijuana at the national level will contribute to the reduction of potential users in the general population. How *feasible* is this policy option?	Definitely Feasible 1 [ ]	Possibly Feasible 2 [ ]	Possibly Unfeasible 3 [ ]	Definitely Unfeasible 4 [ ]	No Judgment 0 [ ]

Source: Items adapted from Jillson (2002): 119–154. Data from Johnston et al. (2016).

The calculation and subsequent presentation of these summary measures of central tendency, dispersion, and polarization are best illustrated graphically. Assume for purposes of illustration that ten advocates in the first round of a hypothetical policy Delphi provided different assessments of the desirability and feasibility of two drug-control goals: to reduce the supply of illicit drugs and to increase public awareness of the difference between responsible and irresponsible drug use. Let us imagine that the responses were those presented in Table 4.9.

Observe that some respondents (advocates 2, 8, and 9) believe that the goal of reducing the supply of illicit drugs is very undesirable but possibly or definitely feasible, whereas others (advocates 1, 5, 7, and 10) believe that this goal is very desirable but definitely unfeasible. When we compare these inconsistencies between desirability and feasibility with responses under goal 2 (public awareness), we find much less inconsistency in the latter set of scores. All of this suggests that the responses to goal 1, although much lower in average desirability and feasibility, also reflect the kinds of important conflicts that

TABLE 4.9

Hypothetical Responses in First-Round Policy Delphi: Desirability and Feasibility of Drug-Control Objectives

	Goal 1 (Reduce Supply)		Goal 2 (Public Awareness)	
Respondent	*Desirability*	*Feasibility*	*Desirability*	*Feasibility*
1	1	4	1	1
2	4	1	2	2
3	3	3	2	1
4	4	2	1	2
5	1	4	2	1
6	2	3	2	1
7	1	4	1	1
8	4	2	1	2
9	4	1	2	2
10	1	4	1	2
	Σ = 25	Σ = 28	Σ = 15	Σ = 15
	Md = 2.5	Md = 3.0	Md = 1.5	Md = 1.5
	Md = 2.5	Md = 2.8	Md = 1.5	Md = 1.5
	Range = 3.0	Range = 3.0	Range = 1.0	Range = 1.0

Note: The median (Md) in a set of scores is the value of the score that falls in the middle when the scores are arranged in order of magnitude. If the number of scores is even, the median is the value of the score that is halfway between the two middle scores. The median is normally used in place of the mean (Mn) when we do not know if the intervals between measures (e.g., the intervals between 1 and 2 and 3 and 4) are equidistant.

policy Delphi is designed to elicit. We would want to report these conflicts as part of the second-round instrument, requesting that respondents provide the reasons and evidence that led them to positions that were so different. Another way to bring out such disagreements is to construct and report an average *polarization measure,* which may be defined as the absolute difference among scores for all combinations of respondents answering a particular question.[55]

Step 5: *Development of subsequent questionnaires.* Questionnaires must be developed for second, third, fourth, or fifth rounds (most policy Delphis involve two to three rounds). As indicated previously, the results of prior rounds are used as a basis for subsequent ones. One of the most important aspects of policy Delphi occurs in these rounds, because it is here that advocates have an opportunity to observe the results of immediately preceding rounds and offer explicit reasons for their respective judgments. Note that later rounds do not simply include information about central tendency, dispersion, and polarization; they also include a summary of arguments offered for the most important conflicting judgments. In this way, the policy Delphi promotes a reasoned debate and maximizes the probability that deviant judgments are not lost in the process. By the time the last round of questionnaires is completed, all have had an opportunity to state their initial positions on forecasts, issues, goals, and options; to examine and evaluate the reasons why their positions differ from those of others; and to reevaluate and change their positions.

Step 6: *Organization of group meeting.* One of the last tasks is to bring advocates together for a face-to-face discussion of the reasons that underlie their various positions. This face-to-face meeting, because it occurs after all advocates have had a chance to reflect on their positions and those of others, may create an atmosphere of informed confidence that would not be possible in a typical committee setting. Face-to-face discussions also create conditions in which advocates may passionately argue their positions and receive immediate feedback.

Step 7: *Preparation of final report.* There is no guarantee that respondents will have reached consensus but much reason to hope that creative ideas about issues, goals, options, and their consequences will be the most important product of a policy Delphi. The report of final results will therefore include a review of the various issues and options available, taking care that all conflicting positions and underlying arguments are presented fully. This report may then be passed on to policymakers, who may use the results of the policy Delphi as one source of information in arriving at decisions.

[55]Ibid. p. 92; and Jerry B. Schneider, "The Policy Delphi: A Regional Planning Application," *Technological Forecasting and Social Change* 3, 4 (1972). The total number of combinations (C) is calculated by the formula $C = k(k - 1)/2$, where k is the number of responses to a particular item. The average difference is obtained by computing the numerical distance between all combinations, adding them up, and dividing by the number of respondents. This requires that we ignore the signs (plus or minus). Another procedure is to retain the signs, square each difference (eliminating minus signs), sum, and average.

Cross-Impact Analysis

Delphi technique is closely related to another widely used judgmental forecasting technique called cross-impact analysis. Developed by the same RAND Corporation researchers responsible for early applications of conventional Delphi,[56] *cross-impact analysis* is a technique that elicits informed judgments about the probability of occurrence of future events on the basis of the occurrence or nonoccurrence of related events. The aim of cross-impact analysis is to identify events that will facilitate or inhibit the occurrence of other related events. Cross-impact analysis was expressly designed as a supplement to conventional Delphi. In the words of two of its early developers,

> A shortcoming of [Delphi] and many other forecasting methods . . . is that potential relationships between the forecasted events may be ignored and the forecasts might well contain mutually reinforcing or mutually exclusive items. [Cross-impact analysis] is an attempt to develop a method by which the probabilities of an item in a forecasted set can be adjusted in view of judgments relating to the potential interactions of the forecasted items.[57]

The basic analytical tool used in cross-impact analysis is the *cross-impact matrix,* a symmetrical table that lists potentially related events along row and column headings (Table 4.10). The type of forecasting problem for which cross-impact analysis is particularly appropriate is one involving a series of interdependent events. Table 4.10, for example, expresses the interdependencies among events that follow the mass production of automobiles. Observe the string of direct positive effects (represented by + signs) directly above the blank cells in the main diagonal. These effects are first-, second-, third-, fourth-, fifth-, and sixth-order impacts of mass automobile production. Note also the positive feedback effects of ($E_2 - E_1$), ($E_3 - E_1$), ($E_7 - E_4$), and ($E_7 - E_5$). These positive feedback effects suggest that ease of travel and patronization of large suburban stores may themselves affect the mass production of automobiles, for example, by increasing demand. Similarly, various forms of social deviance may intensify existing levels of alienation from neighbors and create even greater social-psychological dependence on family members.

This illustration purposefully oversimplifies linkages among events. In many other situations, the linkage of one event with another is not unambiguously positive; nor do events follow one another so neatly in time. Moreover, many events may be negatively linked. For this reason, cross-impact analysis takes into account three aspects of any linkage:

1. *Mode (direction) of linkage.* This indicates whether one event affects the occurrence of another event and, if so, whether the direction of this effect is positive or negative. Positive effects occur in what is called the *enhancing mode,* whereas negative ones fall into a category called the *inhibiting mode.* A good example of linkages in the enhancing mode is increased gasoline prices provoking research and development on synthetic fuels. The arms race and its effects on the availability of funds for urban

[56]These researchers include Olaf Helmer, T. J. Gordon, and H. Hayward. Helmer is credited with coining the term *cross-impact.* See T. J. Gordon and H. Hayward, "Initial Experiments with the Cross-Impact Matrix Method of Forecasting," *Futures* 1, 2 (1968): 101–116.
[57]Ibid., p. 101.

TABLE 4.10

Cross-Impact Matrix Illustrating Consequences of Mass Automobile Use

		Events (E)						
		$E1$	$E2$	$E3$	$E4$	$E5$	$E6$	$E7$
Events (E)	E_1		+	0	0	0	0	0
	E_2			+	0	0	0	0
	E_3		0		+	0	0	0
	E_4	0	0	0		+	0	0
	E_5	0	0	0	0		+	0
	E_6	0	0	0	0	0		+
	E_7	0	0	0			0	

E_1 = mass production of automobiles
E_2 = ease of travel
E_3 = patronization of large suburban stores
E_4 = alienation from neighbors
E_5 = high social-psychological dependence on immediate family members
E_6 = inability of family members to meet mutual social-psychological demands
E_7 = social deviance in form of divorce, alcoholism, juvenile delinquency
Note: A plus (+) indicates direct one-way effects; a zero (0) indicates no effect; a circled plus sign indicates positive feedback effects.

Source: Adapted from Coates (1971).

redevelopment are an illustration of linkages in the inhibiting mode. The *unconnected mode* refers to unconnected events.

2. *Strength of linkage.* This indicates how strongly events are linked, whether in the enhancing or inhibiting mode. Some events are strongly linked, meaning that the occurrence of one event substantially changes the likelihood of another's occurring, whereas other events are weakly linked. In general, the weaker the linkage, the closer it comes to the unconnected mode.
3. *Elapsed time of linkage.* This indicates the amount of time (weeks, years, decades) between the occurrence of linked events. Although events may be strongly linked, either in the enhancing or inhibiting modes, the impact of one event on the other may require a considerable period of time. For example, the linkage between the mass production of automobiles and social deviance required an elapsed time of several decades.

Cross-impact analysis works on the principle of conditional probability. Conditional probability states that the probability of occurrence of one event is dependent on the occurrence of some other event, that is, the two events are not independent. Conditional probabilities may be denoted by $P(E_1|E_2)$, which is read "The probability of the first event (E_1), given the second event (E_2)." For example, the probability (P) of being elected president (E_1) after a candidate has received the party nomination (E_2) may be 0.5; that is, there is a fifty-fifty chance of winning the election [$P(E_1|E_2) = 0.50$]. However, the probability (P) of being elected (E_1) without the party's nomination (E_3) is low, because party nomination is almost a prerequisite for the presidency [$P(E_1|E_3) = 0.20$].

This same logic is extended to cross-impact analysis. The construction of a cross-impact matrix begins with the question: "What is the probability that a certain event (E) will occur prior to some specified point in time?" For example, an extrapolative forecast using time-series analysis may provide an interval estimate that claims that there is a 90 percent (0.9) probability that total energy consumption will exceed 100.0 quadrillion BTUs by 1995. The next question is "What is the probability that this event (E_2) will occur, given that another event (E_1) is certain to precede it?" For example, if presently unpredictable factors (new agreements, political turmoil, accidents) result in a doubling of oil prices (E_1) by 1995, the probability of energy consumption at the level of 100.0 quadrillion BTUs may be reduced to 0.5. In this case, note that "objective" data used to make the original extrapolative forecast are combined with a "subjective" judgment about the conditional probability of the originally projected event, given the prior occurrence of the first.

The construction of a cross-impact matrix for any reasonably complex problem involves many thousands of calculations and requires a computer. Many applications of cross-impact analysis in areas of science and technology policy, environmental policy, transportation policy, and energy policy have involved more than 1,000 separate iterations (called "games" or "plays") to determine the consistency of the cross-impact matrix, that is, to make sure that every sequence of conditional probabilities has been taken into account before a final probability is calculated for each event. Despite the technical complexity of cross-impact analysis, the basic logic of the technique may be readily grasped by considering a simple illustration.

Suppose that a panel of experts assembled for a conventional Delphi provides estimates of the probability of occurrence of four events ($E_1 \ldots E_4$) for future years. Suppose further that these four events are an increase in the price of gasoline to $5 per gallon ($E_1$), the "gentrification" of central city neighborhoods by former suburbanites (E_2), a doubling of reported crimes per capita (E_3), and the mass production of short-distance, battery-powered autos (E_4). The probabilities attached to these four events are, respectively, $P_1 = 0.5$, $P_2 = 0.5$, $P_3 = 0.6$, and $P_4 = 0.2$. Given these subjective estimates, the forecasting problem is this: Given that one of these events occurs (i.e., $P = 1.0$, or 100%), how will the probabilities of other events change?[58]

Table 4.11 illustrates the first round (play) in the construction of a cross-impact matrix. Observe that the assumption that gasoline will go to $5 per gallon produces revised subjective probabilities for "gentrification" (an increase from 0.5 to 0.7), reported crimes per capita (an increase from 0.6 to 0.8), and the mass manufacture of electric autos (an increase from 0.2 to 0.5). These changes reflect the *enhancing linkages* discussed earlier. By contrast, other linkages are *inhibiting*. For example, the gentrification of central cities makes it less likely that gasoline will increase to $5 per gallon (note the decrease from 0.5 to 0.4), on the assumption that oil companies will become more price competitive when former suburbanites drive less. Finally, there are also *unconnected linkages*. Increased crime and mass-produced electric autos exert no influence on the probability of gas prices increasing

[58] Alternatively, we can ask the same question about the *nonoccurrence* of an event. Therefore, it is often necessary to construct two matrices: one for occurrences and one for nonoccurrences.

TABLE 4.11

Hypothetical Illustration of the First Round (Play) in a Cross-Impact Matrix

If This Event Occurs (P = 1.0)	Then the Changed Probability of Occurrence of These Events Is E_1	E_2	E_3	E_4
E_1 Gas to $5 per gallon		0.7	0.8	0.5
E_2 Gentrification	0.4		0.7	0.4
E_3 Crime doubles	0.5	0.4		0.1
E_4 Electric autos	0.4	0.5	0.7	

Events	*Original Probabilities (P)*
E_1	$P_1 = 0.5$
E_2	$P_2 = 0.5$
E_3	$P_3 = 0.6$
E_4	$P_4 = 0.2$

to $5 per gallon or on the process of gentrification (note that original probabilities remain constant at 0.5).

The advantage of the cross-impact matrix is that it enables the analyst to discern interdependencies that otherwise may have gone unnoticed. Cross-impact analysis also permits the continuous revision of prior probabilities on the basis of new assumptions or evidence. If new empirical data become available for some of the events (e.g., crime rates), the matrix may be recalculated. Alternatively, different assumptions may be introduced—perhaps as a consequence of a policy Delphi that yields conflicting estimates and arguments—to determine how sensitive certain events are to changes in other events. Finally, information in a cross-impact matrix may be readily summarized at any point in the process.

Cross-impact matrices may be used to uncover and analyze those complex interdependencies that we have described as ill-structured problems. The technique is also consistent with a variety of related approaches to expert forecasting, including technology assessment, social impact assessment, and technological forecasting.[59] As already noted, cross-impact analysis is not only consistent with conventional Delphi but actually represents its adaptation and natural extension. For example, although cross-impact analysis may be done by single analysts, the accuracy of subjective judgments may be increased by using Delphi panels.

Cross-impact analysis, like the other forecasting techniques discussed in this chapter, has its limitations. First, the analyst can never be sure that all potentially interdependent

[59]On these related approaches, see François Hetman, *Society and the Assessment of Technology* (Paris: Organisation for Economic Co-operation and Development, 1973); Finsterbusch and Wolf, eds., *Methodology of Social Impact Assessment*; and H. A. Linstone, J. F. Coates, T. J. Gordon, M. Adelson, I. Mitroff, D. N. Michael, and M. Turoff, *Forecasting, A New Agenda* (Amsterdam: North-Holland, 1989).

events have been included in the analysis, which again calls attention to the importance of problem structuring (Chapter 3). There are other techniques to assist in identifying these events, including variations of theory mapping (see Table 4.8) and the construction and graphic presentation of networks of causally related events called *relevance trees* (see Chapter 3, Figure 3.14). Second, the construction and "playing" of a cross-impact matrix may be costly and time-consuming, even with the advent of packaged computer programs and high-performance computers. Third, there are technical difficulties associated with matrix calculations (e.g., nonoccurrences are not always analyzed), although many of these problems have been largely resolved.[60]

Feasibility Forecasting

The final judgmental forecasting procedure we consider in this chapter is one that is expressly designed to predict the future behavior of policy stakeholders. This procedure, most simply described as *feasibility forecasting*, assists in producing forecasts about the probable impact of stakeholders in supporting or opposing the adoption and/or implementation of different policies.[61] Feasibility forecasting is particularly well suited for problems requiring estimates of the likely consequences of attempting to implement policy alternatives under conditions of political conflict and the unequal distribution of power and other resources.

Feasibility forecasting may be used to make judgments about the behavior of stakeholders in any phase of the policymaking process, including policy adoption and implementation. What makes this method so useful is that it responds to a key problem that we have already encountered in reviewing other expert forecasting techniques: Typically, there is no relevant theory or available empirical data that permit us to make predictions or projections about the behavior of policy stakeholders. Although social scientists have proposed various theories of policymaking behavior that are potentially available as a source of predictions, most of these theories are insufficiently concrete to apply in specific contexts.[62]

The feasibility assessment technique is one way to respond to a number of concerns about the lack of attention to questions of political feasibility and policy implementation in policy analysis. Although problems of policy implementation play a large part in most policy problems, much of contemporary policy analysis pays little attention to this question. What is needed is a systematic way to forecast "the capabilities, interests, and incentives of organizations to implement each alternative."[63] In practical terms, this means that the behavior of relevant stakeholders must be forecasted along with the consequences of policies

[60]See Dalby, "Practical Refinements to Cross-Impact Matrix Technique," pp. 265–273.

[61]The following discussion draws on Michael K. O'Leary and William D. Coplin, "Teaching Political Strategy Skills with 'The Prince,'" *Policy Analysis* 2, 1 (Winter 1976): 145–160. See also O'Leary and Coplin, *Everyman's "Prince."*

[62]These theories deal with elites, groups, coalitions, and aggregates of individuals and leaders. See, for example, Raymond A. Bauer and Kenneth J. Gergen, eds., *The Study of Policy Formation* (New York: Free Press, 1968), and Daniel A. Mazmanian and Paul A. Sabatier, *Implementation and Public Policy* (Lanham, MD: University Press of America, 1989).

[63]Graham T. Allison, "Implementation Analysis: 'The Missing Chapter' in Conventional Analysis: A Teaching Exercise," in *Benefit-Cost and Policy Analysis: 1974,* ed. Richard Zeckhauser (Chicago, IL: Aldine Publishing Company, 1975), p. 379.

themselves. Only in this way can the analyst ensure that organizational and political factors that may be critical to the adoption and implementation of a policy are adequately accounted for.

Feasibility forecasting, like other judgmental forecasting methods, is based on informed subjective judgments. The method may be used by single analysts, or by a group, much as in a Delphi exercise. Feasibility forecasting focuses on several aspects of political and organizational behavior:

1. *Issue position.* The analyst estimates the probability that various stakeholders will support, oppose, or be indifferent to each of two or more policy alternatives. Positions are coded as supporting (+1), opposing (–1), or indifferent (0). A subjective estimate is then made of the probability that each stakeholder will adopt the coded position. This estimate (which ranges from 0 to 1.0) indicates the saliency or importance of the issue to each stakeholder.
2. *Available resources.* The analyst provides a subjective estimate of the resources available to each of the stakeholders in pursuing his or her respective position. Available resources include prestige, legitimacy, budget, staff, and access to information and communications. Because stakeholders nearly always have positions on other issues for which part of their resources are necessary, available resources may be stated as a fraction or proportion of total resources held by the stakeholder. The resource availability scale, expressed as a fraction, varies from 0 to 1.0. Note that there may be a high probability that a given stakeholder will support a policy (e.g., 0.9), yet that same stakeholder may have little capability to affect the policy's adoption or implementation. This is typically the result of an over-commitment of resources (prestige, budget, staff) to other issues.
3. *Relative resource rank.* The analyst determines the relative rank of each stakeholder with respect to his or her resources. Relative resource rank, one measure of the "power" or "influence" of stakeholders, provides information about the magnitude of political and organizational resources available to each stakeholder. A stakeholder who commits a high fraction (say, 0.8) of available resources to support a policy may nevertheless be unable to affect significantly a policy's adoption or implementation because of inadequate resources.

Because one purpose of feasibility forecasting is to foresee behavior under conditions of political competition and conflict, it is essential to identify as representative and powerful a sample of stakeholders as possible. The analyst may identify representative stakeholders from various organizations and organizational levels with different constituencies and with varying levels of resources and roles in the policymaking process.

An illustration of the feasibility assessment technique appears in Table 4.12. In this example, a policy analyst in a large municipality has completed a study that shows that local property taxes must be raised by 1 percent in order to cover expenditures for the next year. Alternatively, expenditures for municipal services must be cut by a comparable amount, an action that will result in the firing of many employees. Because most public employees are unionized, and there has been a threat of a strike for more than 2 years, the mayor is hesitant to press for a cutback in services. At the same time, local taxpayers' groups are strongly opposed to a tax increase, even if it means the loss of some services. Under these conditions, the mayor has requested a feasibility forecast of the policy alternatives.

Table 4.12 shows that a tax increase is not feasible. The negative sign of the indices (simple and adjusted) at the bottom of the table indicates that there is more opposition than support for a tax increase. By contrast, the index of total feasibility (adjusted) for the budget cut is positive and in the weak range. The index of total feasibility ranges from −1.0

TABLE 4.12

Feasibility Assessment of Two Fiscal Policy Alternatives

(a) Alternative 1 (Tax Increase)

Stakeholder (1)	*Coded Position (2)*	*Probability (3)*	*Fraction of Resources Available (4)*	*Resource Rank (5)*	*Feasibility Score (6)*
Mayor	+1	0.2	0.2	0.4	0.016
Council	−1	0.6	0.7	0.8	−0.336
Taxpayers' association	−1	0.9	0.8	1.0	−0.720
Employees' union	+1	0.9	0.6	0.6	0.324
Mass media	+1	0.1	0.5	0.2	0.010
					ΣF = −0.706

Adjusted total feasibility $(TF_{ADJ}) = TF(5/2) = -0.14(2.5) = -0.35$

(b) Alternative 2 (Budget Cut)

Stakeholder (1)	*Coded Position (2)*	*Probability (3)*	*Fraction of Resources Available (4)*	*Resource Rank (5)*	*Feasibility Score (6)*
Mayor	+1	0.8	0.2	0.4	0.192
Council	+1	0.4	0.5	0.8	0.160
Taxpayers' Association	+1	0.9	0.7	1.0	0.630
Employees' Union	−1	0.9	0.8	0.6	−0.432
Mass Media	−1	0.1	0.5	0.2	−0.010
					$\Sigma F = 0.54$

Index if total feasibility $(TF) = \frac{\Sigma F}{n} = \frac{0.54}{5} = 0.11$

Adjusted total feasibility $(TF_{ADJ}) TF(5/3) = 0.11(5/3) = 0.11(1.66) = 0.183$

to +1.0, which means that the feasibility of different policy alternatives is directly comparable. Observe, however, that there is an adjusted total feasibility index. The adjustment is necessary because the maximum value that the index can take depends on the number of positive (or negative) positions taken.

In Table 4.12(a), for example, the maximum negative value of the index depends on the number of negative positions in relation to positive ones. Imagine that the two negative feasibility scores were "perfect" (i.e., –1.0) and that all other stakeholders were indifferent, giving them feasibility scores of 0.0. In this case, the maximum value of the sum of individual feasibility scores would be –2.0. If we then divide by the total number of stakeholders ($n = 5$), we produce an index value of –0.40, although there is maximum opposition to the alternative. For this reason, we must compute a maximum value of TF, which in this case is $TF_{MAX} = 2/5 = 0.40$. To find the adjusted value of the index (TF_{ADJ}), we simply divide the original value of TF by its maximum value, that is, $TF/TF_{MAX} = TF_{ADJ} = -\,0.40/0.40 = -1.0$. The same procedure was used to find the maximum and adjusted positive values of TF in Table 4.12(b).

Feasibility forecasting forces us to make explicit subjective judgments, rather than treat political and organizational questions in a loose or arbitrary fashion. Feasibility assessment also enables analysts to systematically consider the sensitivity of issue positions and available resources to changes in policy alternatives. In the previous example, a smaller tax increase, combined with austerity measures and a plan to increase the productivity of municipal employees, might evoke an altogether different level of feasibility.

Feasibility forecasting, like conventional Delphi and cross-impact analysis, provides no systematic way of surfacing the reasons and evidence that underlie subjective judgments. Perhaps the best way to resolve this difficulty is to adopt procedures from policy Delphi or use assumptional analysis (Chapter 3). A second limitation is that it assumes that the positions of stakeholders are independent and that they occur at the same point in time. These assumptions are unrealistic, because they ignore processes of coalition formation over time and the fact that one stakeholder's position is frequently determined by changes in the position of another. To capture the interdependencies of issue positions as they shift over time, we might employ some adaptation of cross-impact analysis. Finally, feasibility forecasting, like other judgmental forecasting techniques, is most useful under conditions in which the complexity of a problem cannot easily be grasped by using available theories or empirical data. For this reason, any attempt to outline its limitations should also consider its potential as a source of creative insight.

In concluding this chapter, it is important to stress that improvements in forecasting are likely to result from the creative combination of different approaches and techniques, that is, from *multimethod forecasting.* Multimethod forecasting combines multiple forms of logical reasoning (inductive, deductive, abductive) and multiple bases (extrapolation, theory, judgment). Multimethod forecasting recognizes that neither precision nor creativity is an end in itself. What appears to be a creative or insightful may lack plausibility and turn out to be pure speculation or quackery, whereas highly precise projections or predictions may simply answer the wrong question. The ultimate justification for a forecast is whether it provides plausibly true beliefs about the future.

EVIDENCE-BASED FORECASTING

Throughout this book, we stress the importance of achieving high standards of evidence-based policy analysis. In this context, it is important to consider evidence-based forecasting as one aspect of meeting these standards, considering the integral nature of forecasting in policy analysis today. J. Scott Armstrong and his colleagues have been engaged in a major successful effort to compile on the basis of the experience of real-world forecasters principles of evidence-based forecasting.[64] These principles have been organized in several important areas: problem formulation, information search, choice of methods, evaluation of forecasts, and the use of forecasts by policymakers and planners.

Although Armstrong lists more than 150 criteria for developing evidence-based forecasts, we will consider only the most directly relevant to our aims here (Table 4.13). The evidence-based criteria are stated here as recommendations for the conduct of extrapolative, theoretical, and judgmental forecasts.

TABLE 4.13

Recommendations for Evidence-Based Forecasting

RECOMMENDATION	PURPOSE
Describe decisions that might be affected by the forecast.	Maximize the relevance of the forecast to policymakers.
Decompose the problem into subproblems.	Ensure that the problem has been adequately structured.
Use theory to identify explanatory variables.	Maximize the likelihood of identifying a plausible causal mechanism.
Assess the reliability and validity of questionnaires and scales.	Avoid possibility of error that is artificially produced by null findings.
Use structured rather that unstructured forecasting methods.	Minimize unknown sources of systematic bias.
Pretest instruments prior to the conduct of the survey.	Discover errors prior to using instruments.
Obtain forecasts from a heterogeneous sample of experts.	Maximize the availability of multiple perspectives.
Estimate confidence intervals (CIs).	Express point estimates probabilistically.
Provide clear explanation of forecasting methods.	Ensure that experts are competent to answer.

[64]J. Scott Armstrong, "Findings from Evidence-Based Forecasting: Methods for Reducing Forecast Error." *International Journal of Forecasting* 22, 3 (2006): 583–598; and J. S. Armstrong (Ed.) *Principles of Forecasting: A Handbook for Researchers and Practitioners.* Springer Science & Business Media. See also Armstrong's website, www.forecastingprinciples.com.

CHAPTER SUMMARY

This chapter has provided an overview of the process of forecasting, highlighting the nature, types, and uses of forecasting in policy analysis. After comparing approaches based on inductive, deductive, and abductive reasoning, specific methods and techniques were described. These included methods and techniques of extrapolative, theoretical, and judgmental forecasting. The ultimate justification for a forecast is whether it yields plausibly true beliefs about the future, not whether it is based on a particular method, quantitative or qualitative. In the words of an accomplished policy analyst and former assistant secretary in the Department of Defense, "it is better to be roughly right than exactly wrong."[65]

REVIEW QUESTIONS

1. What are the three forms of forecasting and how are they related to bases of forecasts?
2. In addition to promoting greater understanding of the future, what other aims are achieved through forecasting?
3. To what extent is econometric forecasting more accurate than methods of extrapolative forecasting and forecasting with expert judgment? Explain.
4. How do the institutional, temporal, and historical contexts of forecasts affect their accuracy?
5. Distinguish among potential, plausible, and normative futures.
6. Contrast inductive, deductive, and abductive (retroductive) reasoning. Do the same for theoretical and judgmental forecasting.
7. List and describe techniques used in the three main types of forecasts. Provide examples.
8. Whether they are linear or nonlinear, most forecasts are based on assumptions of persistence and regularity. To what extent are these assumptions plausible?
9. Many forecasts employ linear regression analysis, also known as the classical linear regression (CLR) model. What are the main assumptions of the CLR model when used to make forecasts?
10. What corrective actions can be taken when assumptions of the CLR model are violated?
11. Is it better to be approximately right, rather than exactly wrong? How does your answer affect the choice of a forecasting method?
12. Important examples of multimethod forecasting may be found in the work conducted by the Good Judgment Project. Describe the method of conducting prediction markets and how this method compares with forecasts based on expert judgment (https://www.iarpa.gov/index.php/newsroom/iarpa-in-the-news/2015/439-the-good-judgment-project).
13. There is an emerging consensus that tests of significance hamper the interpretation of results of forecasts as well as other methods, for example, the analysis of policy outcomes as effects of programs. Why? (T. Ziliak and D. McCloskey, *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives*. University of Michigan Press, 2008).

[65] Alain C. Enthoven, "Ten Practical Principles for Policy and Program Analysis," in *Benefit-Cost and Policy Analysis: 1974*, ed. Zeckhauser, p. 459.

DEMONSTRATION EXERCISES

1. After reading Case 4.1, answer these questions:
   - Explain why the r, r^2, and p-values are identical in Figure C4.1.1(a) through (d).
   - Describe the patterns in each scatterplot (a) through (d).
   - Explain the importance of inspecting a visual signature of a regression analysis.
2. After looking at the data on receipts in Case 4.2, use SPSS or a similar statistical program (Excel, Stata, or R) to perform a forecast in which you estimate the value of Metropolitan Atlanta Regional Transit Authority (MARTA) receipts in the year 2004. Assume you are conducting the analysis in January 2000, as a consultant to MARTA. The client has hired you because MARTA needs an accurate estimate of future receipts. The estimate will be used to make decisions about fare increases that have become a focal point of debates about environmental justice, that is, the justice of the burden of paying for transportation among minorities and the poor. Although you did not know it at the time, the Atlanta City Council opposed a fare increase in 2000 and 2001, amidst considerable political turmoil. MARTA's expenditures in 2001 were \$307 million, while revenues were approximately \$304 million, a shortfall that compelled MARTA policymakers to increase fares. Community groups contended that the fare increase violated principles of environmental justice.

   Perform the following analyses, which are based on the procedures for extrapolative forecasting covered in the chapter:
   - Enter the data from Table C4.2.1 (Payments to MARTA, 1973–2014) into the SPSS (or other) Data Editor. Note that the value of RECEIPTS is in thousands of dollars.
   - Create a sequence (time-series) graph of the variable RECEIPTS. Does the series meet assumptions of the linear regression model?
   - Perform a logarithmic transportation and create a new sequence graph. What difference occurs as a result of the log transformation?
   - Using data on receipts from 1973–2000, perform a time-series regression analysis to forecast the value of receipts in 2004, 2008, and 2014. This will require three regression analyses. Compute a 95 percent confidence interval for the mean. What are the point estimates for the 3 years.
   - How do the short-, medium-, and longer-term forecasts compare with the actual values of receipts for the 3 years? Do the point estimates fall within the 95 percent confidence intervals?
   - Interpret the point and interval estimates provided in the output. How accurate is your estimate for the 3 years? Does MARTA need to raise fares? How sure are you?
   - Write a 2-page policy memo to MARTA in which you explain and justify your forecast for the 3 years. Include a policy recommendation about options for increasing fares and relate your recommendation to the issue of environmental justice.
3. After reading Case 4.3, explain how the steps in a Delphi forecast presented in this chapter are related to the "wisdom of the crowd," as represented by the line graphs in Figure C4.3.1

BIBLIOGRAPHY

Allen, T. Harrell. *New Methods in Social Science Research: Policy Sciences and Futures Research.* New York: Frederick A. Praeger, 1978.

Anscombe, F. J. "Graphs in Statistical Analysis." *American Statistician* 27 (February 1973): 17–21.

Armstrong, J. Scott. "Findings from Evidence-Based Forecasting: Methods for Reducing Forecast Error." *International Journal of Forecasting* 22, 3 (2006): 583–598.

Armstrong, J. Scott. *Illusions in Regression Analysis* (2011). Available at SSRN: https://ssrn.com/abstract=1969740.

Armstrong, J. Scott, *Long-Range Forecasting*, 2nd ed. (August 11, 2010). Available at SSRN: https://ssrn.com/abstract=666990.

Armstrong, J. Scott, ed. *Principles of Forecasting: A Handbook for Researchers and Practitioners*. New York: Springer Science & Business Media, 2001.

Armstrong, J. Scott. "Significance Tests Harm Progress in Forecasting." *International Journal of Forecasting* 23, 2 (2007): 321–327.

Ascher, William. *Forecasting: An Appraisal for Policy Makers and Planners.* Baltimore, MD: Johns Hopkins University Press, 1978.

Ascher, William. "The Forecasting Potential of Complex Models." *Policy Sciences* 13 (1981): 247–267.

Box, G.E.P., and G. M. Jenkins. *Time Series Analysis: Forecasting and Control.* San Francisco, CA: Holden-Day, 1969.

Campbell, Mary P. "What I've Learned from the Good Judgment Project." *Forecasting and Futurism* 11 (July 2015): 18–22.

Coates, Joseph. "Technology Assessment: The Benefits, the Costs, the Consequences." *Futurist 5,* 6 (December 1971): 219–225.

Finsterbusch, Kurt, and C. P. Wolf, eds. *Methodology of Social Impact Assessment.* Stroudsburg, PA: Dowden, Hutchinson & Ross, 1977.

Gass, Saul I., and Roger L. Sisson, eds. *A Guide to Models in Governmental Planning and Operations.* Washington, DC: U.S. Environmental Protection Agency, 1974.

Guess, George M., and Paul G. Farnham. *Cases in Public Policy Analysis.* New York: Longman, 1989, Ch. 3, "Forecasting Policy Options," pp. 49–67.

Gurr, Ted R. "The Comparative Analysis of Public Order," in *The Politics of Crime and Conflict,* ed. Ted R. Gurr, Peter N. Grabosky, and Richard C. Hula (Beverly Hills, CA: Sage Publications, 1977).

Harrison, Daniel P. *Social Forecasting Methodology.* New York: Russell Sage Foundation, 1976.

Helmer, Olaf, and Nicholas Rescher. "On the Epistemology of The Inexact Sciences." *Management Science* 6, 1 (1959): 25–52.

Horowitz, Michael. "Good Judgment in Forecasting International Affairs (and an Invitation for Season 3)." *The Washington Post* (November 26, 2013). https://www.washingtonpost.com/news/monkey-cage/wp/2013/11/26/good-judgment-in-forecasting-international-affairs-and-an-invitation-for-season-3/.

Intelligence Advanced Research Projects Activity (IARPA). *The Good Judgment Project*. https://www.iarpa.gov/index.php/newsroom/iarpa-in-the-news/2015/439-the-good-judgment-project

Jillson, Irene A. "The National Drug Abuse Policy Delphi: Progress Report and Findings to Date," in *The Delphi Method: Techniques and Applications,* ed. Harold A. Linstone and Murray Turoff. New York: Addison-Wesley, 2002.

Johnston, Lloyd D., P. M. O'Malley, R. A. Miech, J. G. Bachman, and J. E. Schulenberg, eds. *Monitoring the Future National Survey Results on Drug Use, 1975–2010.* Volume II, College Students & Adults Ages 19–50. Ann Arbor: University of Michigan, Institute for Social Research, 2016.

Kahneman, Daniel. "A Perspective on Judgement and Choice: Mapping Bounded Rationality." *American Psychologist* 58, 9 (2003): 697.

Kaufman, Herbert. *Are Government Organizations Immortal?* Washington, DC: Brookings Institution, 1976.

Linstone, Harold A., and Murray Turoff, eds. *The Delphi Method: Techniques and Applications.* Reading, MA: Addison-Wesley, 1975.

Marino, Y. P. *Technological Forecasting for Decision Making.* Cambridge, MA: American Elsevier Company, 1972.

McNown, Robert. "On the Use of Econometric Models: A Guide for Policy Makers." *Policy Sciences* 19 (1986): 360–380.

O'Leary, Michael K., and William D. Coplin. *Everyman's "Prince."* North Scituate, MA: Duxbury Press, 1976.

Rescher, Nicholas. *Predicting the Future: An Introduction to the Theory of Forecasting.* Albany, NY: SUNY Press, 1998.

Schroeder, Larry D., and Roy Bahl. "The Role of Multi-Year Forecasting in the Annual Budgeting Process for Local Governments." *Public Budgeting and Finance* 4, 1 (1984): 3–14.

Tetlock, Philip. *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton, NJ: Princeton University Press, 2005.

Thomopoulos, Nick T. *Applied Forecasting Methods.* Englewood Cliffs, NJ: Prentice Hall, 1980.

Toulmin, Llewellyn M., and Glendal E. Wright, "Expenditure Forecasting." In *Handbook on Public Budgeting and Financial Management.* Edited by Jack Rabin and Thomas D. Lynch. New York: Marcel Dekker, 1983, pp. 209–287.

Tufte, Edward R. *The Visual Display of Quantitative Information*. Cheshire, CT: The Graphics Press, 1983.

Ungar, L., B. Mellors, V. Satopää, J. Baron, P. Tetlock, J., Ramos, and S. Swift. *The Good Judgment Project: A Large Scale Test.* AAAI Technical Report FS-12–06. Philadelphia, PA: University of Pennsylvania, 2012.

U.S. General Accounting Office. *Prospective Evaluation Methods: The Prospective Evaluation Synthesis.* Washington, DC: U.S. General Accounting Office, Program Evaluation and Methodology Division, July 1989.

CASE 4.1 VISUAL SIGNATURES IN REGRESSION ANALYSIS: THE ANSCOMBE QUARTET

The four scatterplots in Figure C4.1 show different patterns among the observations on the graph. Each set of observations of X and Y yields identical correlation coefficients (r), coefficients of determination (r2), and p-values.

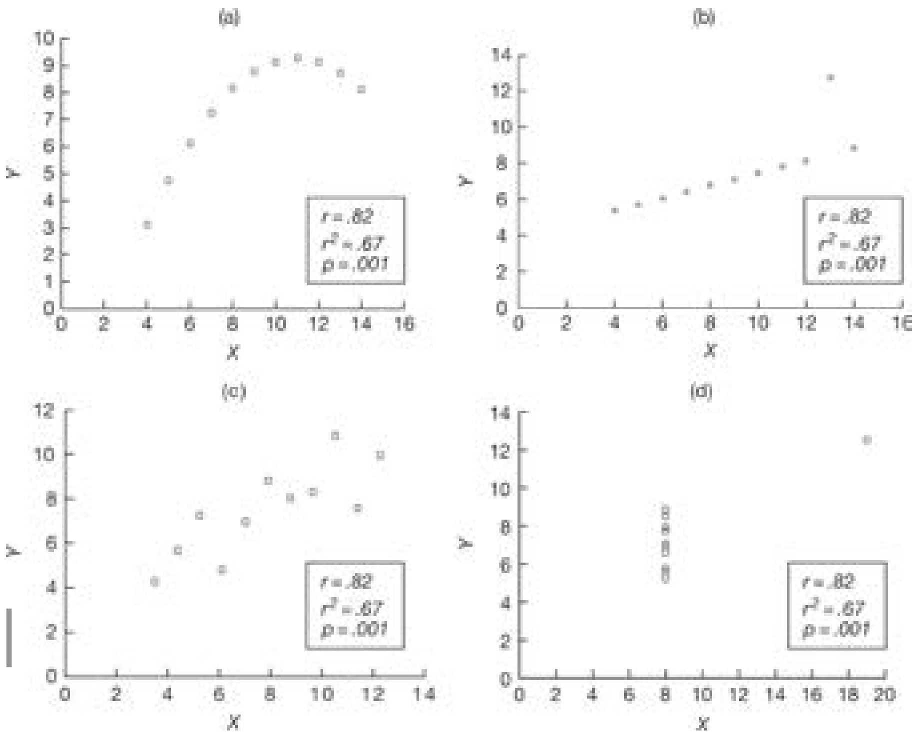


FIGURE C4.1.1
The Anscombe Quartet

CASE 4.2 FORECASTING MARTA RECEIPTS

The following are statistics on MARTA receipts from sales tax and user fees from 1973 to 2010.

TABLE C4.2.1

MARTA Receipts, 1973–2010

Year	Receipts
1973	$43,820
1974	$50,501
1975	$50,946
1976	$52,819
1977	$57,933
1978	$66,120
1979	$75,472
1980	$88,342
1981	$99,836
1982	$104,685
1983	$112,008
1984	$123,407
1985	$134,902
1986	$147,149
1987	$148,582
1988	$158,549
1989	$162,543
1990	$165,722
1991	$168,085
1992	$167,016
1993	$181,345
1994	$198,490
1995	$222,475
1996	$229,510
1997	$236,545
1998	$243,579
1999	$277,309
2000	$299,103
2001	$303,562
2002	$284,427

(*continued*)

2003	$271,006
2004	$283,381
2005	$307,227
2006	$334,486
2007	$350,526
2008	$349,668
2009	$312,704
2010	$307,525

Source: Guess and Farnham (1989), p. 196; and 2010 Comprehensive Annual Financial Report, Metropolitan Atlanta Regional Planning Authority, 2010. Receipts for 1996 and 1997 are linear interpolations.

CASE 4.3 MULTIMETHOD FORECASTING: SYRIA CHEMICAL WEAPONS

In 2005, the business and intelligence forecasting communities were turned upside down by the publication of James Surowiecki's book, *The Wisdom of Crowds* (New York: Anchor Books, 2005). Surowiecki told the story of Sir Francis Galton, the British scientist and inventor of regression analysis, who at a fair in 1906 observed some 800 villagers participating in a weight-judging competition. The winners would receive a reward for correctly guessing the weight of a slaughtered ox.

Galton the elitist, skeptical of the wisdom of the crowd, was surprised to find that the crowd guessed the weight of a slaughtered ox within one pound of its actual weight of 1,198 pounds. Each member of the crowd guessed as individuals, with little or no discussion with neighbors. Like modern Delphi forecasters, they seem to have been essentially anonymous, without being affected by the views of dominant members of the crowd.

The Good Judgment Project, sponsored by the Intelligence Advanced Research Projects Activity

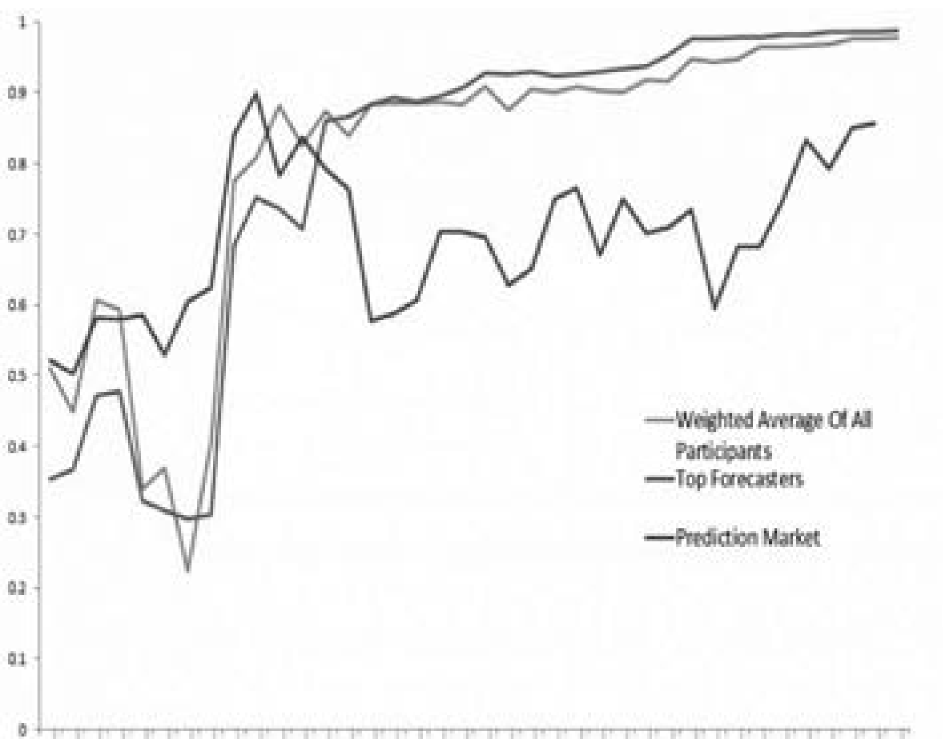


FIGURE C4.3.1
Multimethod Forecast of Chemical Weapons in Syria

Source: Horowitz (2013).

(*continued*)

(IARPA) (https://www.iarpa.gov/) of the U.S. Director of National Intelligence, is based on ideas of the wisdom of crowds. It is also an extension of the seminal work of Philip E. Tetlock, *Expert Political Judgment: How Good Is It? How Can We Know?* (Princeton, NJ: Princeton University Press, 2005), the first systematic and critical analysis of social, political, and economic predictions by expert forecasters.

In early 2013, one of the research initiatives of the Good Judgment Project was designed to compare the accuracy of three methods of forecasting: the weighted judgments of non-expert forecasters; the judgments of top forecasters supplemented by training in forecasting best practices; and the judgments of members of prediction markets.[66] While the top forecasters may be thought of as experts, they were in fact highly motivated amateurs more than they were "experts" in the conventional sense.

Over an 11-month period, the forecasts of the weighted average group and the top forecasters converged, as shown in Figure C4.3.1, with the prediction market group less accurate. The task of the three groups was to place a probability from 0.0 through 1.0 on the likelihood that the Organization for the Prevention of Chemical Weapons (OPCW) would inspect, collect, and destroy chemical weapons in Syria by December 1, 2013. This was a profoundly important and complex policy issue, because it involved President Obama working with President Putin in taking an action that many considered brilliant: The destruction of the chemical weapons would protect Syrian and other citizens, while avoiding the kind of full-scale conflict planned under Obama's congressionally sanctioned "red line" in Syria.

Figure C4.3.1 shows that the weighted judgments of non-experts were as accurate as the top forecasters and more accurate than the prediction market for chemical weapons. The Good Judgment Project demonstrated the ability "to generate increasingly accurate forecasts that have exceeded even some of the most optimistic estimates. . . . Since the question resolved as a "yes," the closer each forecasting method's predictions were to 1, the better they did. As the graph shows, after some initial uncertainty about whether inspections would occur, our forecasters generally converged on the correct answer well before the outcome."[67] Consistent with these results, the literature suggests that structured approaches such as Delphi provide more accurate forecasts than unstructured traditional meetings.[68] ■

[66]The idea of a prediction market (for a time named Policy Analysis Market) is credited to the Austrian economists Friedrich Hayek and Ludwig von Mises. See Hayek, "The Use of Knowledge in Society," *American Economic Review 35*, 4: 519–530. A prediction market is an exchange-traded market created to investigate the outcome of events. The prices of these events, estimated by many persons (the "crowd"), serve as a proxy for the probability of occurrence of the event. Like Delphi, the forecasters make essentially anonymous judgments as individual entrepreneurs in a market.

[67]Throughout 2014 and 2015, the news media reported on the use of chlorine gas, which was used in Anbar (Iraq) and Syria. It is not on the list of banned weapons, primarily because chlorine has many civilian uses. Chlorine gas proved to be ineffective as a weapon in World War I and was replaced by phosphine and mustard gas.

[68]A. Graefe and J. S. Armstrong "Comparing Face-To-Face Meetings, Nominal Groups, Delphi and Prediction Markets on an Estimation Task," *International Journal of Forecasting*, 27, 1: 183–195.

CHAPTER 5

Prescribing Preferred Policies

LEARNING OBJECTIVES

By studying this chapter you should be able to:

- Describe criteria employed to make choices among two or more policy alternatives
- Contrast the Pareto criterion, the Kaldor–Hicks criterion, and the Rawls criterion
- Compare and contrast comprehensive rationality and disjointed incrementalism as models of choice
- Describe the reasons underlying different types of rationality
- Identify steps in conducting a benefit–cost analysis
- Discuss the main sources of uncertainty in making prescriptions
- Describe threats to the plausibility of policy prescriptions
- Use a case study to perform a critique of preliminary benefit–cost analysis

INTRODUCTION

What should be done? Information needed to answer this question is produced by the methodology of prescription. While the methodology of forecasting covered in the last chapter is indispensable for producing information about expected policy outcomes, it does not provide a means for choosing among two or more expected outcomes. Hence, while forecasting answers questions about the expected outcomes of policies, it does not answer questions about which policy to choose. That is the job of prescription.

This chapter begins with an overview of prescription, focusing on policy choice as a fundamental task of policy analysis. Central to policy choice is the consideration of reasons for making *rational* choices, especially choices based on normative economic reasoning including the analysis of opportunity costs incurred when one policy alternative is chosen rather than another. We then proceed to approaches to prescription, the most important of which are cost–benefit and cost–effectiveness analysis. Associated with these approaches are a number of techniques that assist in making choices among policy alternatives whose expected outcomes are known.

PRESCRIPTION IN POLICY ANALYSIS

The methodology of prescription transforms information about expected policy outcomes into information about preferred policies. To prescribe a policy requires prior information about the expected outcomes of choosing among alternative policies. However, the methodology of prescription requires that we also have information about which outcomes are more valuable than others. For this reason, prescription is inherently normative and closely related to ethical and moral issues.[1]

Prescription and Policy Advocacy

Should the United States increase its economic commitment to less-developed countries by raising levels of foreign aid and technical assistance? Should Congress pass legislation that will strictly curtail the pollution of the atmosphere and waterways by industry and by drivers of vehicles used for work or recreation? Should state governments provide low-cost home-heating fuel to those who cannot afford it? Should the city council raise local taxes to build a community recreation center, or maintain a balanced budget? Should the federal government provide a minimum annual income for all citizens or invest in a cure for cancer?

These policy issues call for prescriptions that answer the question: What should be done? Answers to this question call for an approach that is normative, rather than one that

[1]Daniel M. Hausman and Michael S. McPherson, *Economic Analysis, Moral Philosophy and Public Policy*, 2nd ed. (Cambridge: Cambridge University Press, 2006); Dunn, "Values, Ethics, and Standards in Policy Analysis," in *Encyclopedia of Policy Studies*, ed. Stuart S. Nagel (New York: Marcel Dekker, 1983), pp. 831–866; and Stone, *Policy Paradox*, Part I: Goals, pp. 35–130.

is merely empirical, because the question is one of right action. Questions of right action involve advocative claims,[2] which have several characteristics:

1. *Actionable.* Advocative claims focus on actionable alternatives that may be taken to resolve a policy problem. Although advocative claims require prior information about what will occur and what is valuable, they go beyond questions of "fact" and "value" because they include arguments about what should be done.
2. *Prospective.* Advocative claims are prospective, because they occur prior to the time that actions are taken (*ex ante*), while monitoring and evaluation are retrospective, because they are applied after actions are taken (*ex post*). By contrast, forecasting and prescription are both applied prospectively (*ex ante*).
3. *Value laden.* Advocative claims depend as much on "facts" as they do on "values." To claim that a particular policy alternative should be adopted requires not only that the action being prescribed will have the predicted consequences; it also requires that the predicted consequences are valued.[3]
4. *Ethically complex.* The values underlying advocative claims are ethically complex. A given value (e.g., health) may be intrinsic as well as extrinsic. *Intrinsic values* are those that are valued as ends in themselves; *extrinsic values* are those that are valued because they lead to a valued end. For example, health may be regarded as an end in itself and as a means for the attainment of other values, including wealth, enlightenment, security, freedom, and economic efficiency.

Multiple advocacy, which is coterminous with rational advocacy, calls for the systematic comparison and critical assessment of multiple potential solutions; it is not a way to defend a single position at any cost. To be sure, analysts may arrive at a single prescription, but only after critically assessing the pros and cons of multiple prescriptions. Multiple advocacy is as much an approach to problem structuring as it is to problem solving.[4] Multiple advocacy is less likely to lead to an *over-advocacy trap* (Box 5.1), a trap that yields faulty solutions because analysts have defined the wrong problem and the wrong solution.

A Simple Model of Choice

Advocative claims are called for when analysts face a choice between two or more alternatives. In some situations, the choice is between a new course of action and the status quo. In other situations, there may be many alternatives.

In its simplest terms, choice is a process of reasoning involving three interrelated processes: (1) the definition of a problem requiring action, (2) the comparison of consequences

[2]Here we are describing rational advocacy, not policy advocacy. Policy advocacy involves a predetermined choice of a politically favored alternative, without regard to evidence, whereas rational advocacy involves the use of reason and evidence to choose alternatives that may or may not be favored by the analyst or her client.

[3]Prescriptions involve the combination of a value premise and a factual premise.

[4]Ian I. Mitroff, Richard O. Mason, and Vincent P. Barabba, *The 1980 Census: Policymaking amid Turbulence* (Lexington, MA: D. C. Heath, 1983), see the advocacy of two or more policies (multiple advocacy) as a methodology of problem structuring as well as problem solving, as does Alexander George, *Presidential Decision Making in Foreign Policy: The Effective Use of Information and Advice* (Boulder, CO: Westview Press, 1980).

BOX 5.1 THE OVER-ADVOCACY TRAP

In *Presidential Decision Making and Foreign Policy: The Effective Use of Information and Advice* (1980), Alexander George lists shortcomings and weaknesses of policy advice that he describes as the *over-advocacy trap.** The over-advocacy trap occurs when:

- The client and the analyst agree too readily on the nature of the problem and responses to it.
- Disagreements among policy advocates incorporated in an analysis do not cover the full range of policy alternatives.
- The analyst ignores advocates for unpopular policy alternatives.
- The analyst fails to communicate prescriptions requiring that the client face up to a difficult or unpopular decision.
- The analyst is dependent on a single source of information.
- The client is dependent on a single analyst.
- Assumptions of a policy are evaluated only by advocates of that policy—including the trapped analyst or client.
- The client dismisses the results of analysis simply because they are perceived as negative or counterintuitive.
- The client or the analyst uncritically accepts consensus findings without probing the basis for the consensus and how it was achieved. ■

Source: *George (1980): 23–24. I have replaced the term *adviser* with *analyst.*

of two or more alternatives that may resolve the problem, and (3) the prescription of the alternative that will result in the preferred outcome. Choice may be described as a process of reasoning where the first alternative, (A_1), yields one outcome, (O_1), and the second alternative, (A_2), yields another outcome, (O_2). If the value of O_1 is greater than O_2 ($O_1 > O_2$), A_1 is prescribed as the preferred alternative.

This simple process of reasoning contains two essential elements of choice: factual and value premises. The first decision premise states that A_1 will result in O_1 while the second decision premise states that A_2 will result in O_2. These are *factual premises*, that is, beliefs that in principle may be shown to be empirically true or false. The third premise, however, is a *value premise,* that is, a belief that outcome O_1 is more effective, efficient, or just than O_2 on some scale of value. Value premises cannot be proved right or wrong by appealing to factual premises, because questions of value are determined to be right or wrong on the basis of ethical and moral principles.

This simple model has the advantage of pointing out that factual and value premises are present in all choice situations. The disadvantage of the model is that it obscures the complexity of choice. Consider, for example, the conditions that must be present for this model of choice to be valid.[5]

1. *Single decision-maker.* The choice must be confined to a single person. If choices are made by more than one person, there are likely to be conflicting factual and value premises.

[5]Richard Zeckhauser and Elmer Schaefer, "Public Policy and Normative Economic Theory," in *The Study of Policy Formation*, ed. Raymond A. Bauer and Kenneth J. Gergen (New York: Free Press, 1968), p. 28.

2. *Certainty.* The outcomes of choice must be known with certainty. Yet outcomes are seldom if ever known with certainty. For example, apart from the choice of a preferred alternative, there are many uncontrollable factors that enhance or inhibit the occurrence of an outcome.
3. *Immediacy of consequences.* Results of a course of action must occur immediately. In most choice situations, however, outcomes are delayed. Because outcomes do not occur immediately, the values that originally prompted the action change over time.

Imagine now that our simple model of choice involves the issue of whether to provide a minimum wage to unskilled workers. Assume that unskilled workers are now covered by the current (2016) minimum wage of $7.25 per hour. On the basis of an analysis, we conclude that if we maintain the status quo ($7.25 per hour) the result will be an annual average part-time income of $4,000 for unskilled workers. We might also forecast that a minimum wage of $10.00 per hour will result in an annual average part-time income of $7,000. If we assume that more income is better than less income, we will have no difficulty in choosing as follows:

$$A_1 \rightarrow O_1\,(\$4{,}000)$$
$$A_2 \rightarrow O_2\,(\$7{,}000)$$
$$O_2 > O_1$$
$$\therefore A_2$$

This case fails to satisfy the three conditions necessary for the simple model of choice. First, there are multiple decision-makers, not simply one. Many stakeholders—legislators, voters, administrators, employers, workers—affect and are affected by the issue of minimum wages. Each stakeholder brings different factual and value premises to the choice situation, and there will be disagreements about what should be done and why. Legislators may want to maintain the income of unskilled workers, citizen groups may want to reduce the tax burden that will accompany minimum wage legislation, while labor unions may want to increase the income of unskilled workers. Second, the consequences of minimum wage legislation are uncertain. Many factors other than the legislation—the availability of college students as an alternative source of unskilled labor, the average income of local workers, the level of unemployment—may determine whether or not employers will fire some workers as the price of raising the minimum wage. Finally, the consequences of the legislation will occur over time, which means that values may change. If there is a deep economic recession such as that of the period 2008–2011, groups which formerly supported a $7.25 per hour minimum wage may conclude that the increased wage will increase unemployment. In short, the simple model of choice provides a distorted picture of the facts and values involved in setting a minimum wage for unskilled workers, a picture that fails to recognize that, factually, the minimum wage probably has little or no effect on unemployment.[6]

[6]A 1995 meta-analysis of studies of the effects of the minimum wage on employment, after taking into account sampling and publication biases in previous research, concludes that the minimum wage probably has little negative effect and even a positive effect on employment. David Card and Alan B. Krueger, "Time-Series

A Complex Model of Choice

Suppose that we make a more thorough effort to gather information about the effects of minimum wages. In addition to the original alternatives (minimum wages and the status quo), a third alternative is identified. The third alternative, job training, is based on the assumption that the problem is not low wages, per se, but an absence of skills necessary to qualify workers for higher-paying jobs. After adding this third alternative, we might forecast that the current minimum wage (A_1 or \$7.25 per hour) will produce an average annual income of \$18,802 for some 12,000 unskilled workers; the second alternative (A_2 or a \$10.00 minimum wage) will produce an average annual income of \$24,958 for 6,000 unskilled workers, with the remaining 6,000 workers entering the ranks of the unemployed; the third alternative (A_3 or job training) will produce an average annual income of \$18,808 for some 10,000 newly employed workers.

Each alternative will have different consequences. For example, legislators must be attentive to their prospects for reelection. A minimum wage of \$7.25 will result in the loss of seats in districts where labor and welfare rights organizations are powerful. The costs of the program will be passed on to owners of small businesses, who have little influence in the legislature. Finally, new programs in job training will result in a loss of seats in districts with strong opposition to any new tax increase. In any case, the benefits of training are not immediate and will not be felt for several years. The real earning capacity of trainees will increase gradually, reaching a level of \$22,500 for 12,000 workers by 2016. The real incomes of workers receiving the \$10.00 minimum wage will decline as a result of income taxes. Finally, the three alternatives have costs that will be borne unequally by various stakeholders. The first alternative involves no new costs, but the second will require that owners of small businesses pay the additional costs. The third alternative will distribute costs among taxpayers who will foot the bill.

The simple model of choice has quickly become complex. Now there are multiple stakeholders, uncertainty about outcomes, and time. The problem with the complex model of choice is that we cannot arrive at a satisfactory prescription that combines the values of all stakeholders at several points in time.

A situation such as this, where it is impossible to consistently rank alternatives, is *intransitive*. Intransitive choice typically involves multiple conflicting objectives and should be contrasted with situations involving transitive choices. A *transitive* choice is one where alternatives can be consistently ranked: If A_1 is preferable to A_2 in the choice set (A_1, A_2) and A_2 is preferable to A_3 in the choice set (A_2, A_3), then A_1 is preferable to A_3 in the set (A_1, A_3). Transitive choices can be constructed by assigning values to each alternative, so that if A_1 is preferred to A_2 and A_3, it is assigned a higher value. The person making a choice is said to maximize her utility by selecting the alternative

Minimum-Wage Studies: A Meta-Analysis," *The American Economic Review* 85, 2 (1995): 238–243. Recent natural experiments tend to support Card and Krueger's earlier meta-analysis. See Harriet Dueled and Xiangfei Liu, "Estimating More Precise Treatment Effects in Natural and Actual Experiments", IZA Discussion Papers, No. 10055 (Berlin: Institute for Labor (IZA), 2016).

TABLE 5.1

Transitive and Intransitive Preferences

	(a) Transitive Outcome		
Alternative	*2013 Income*	*2015 Income*	*2016 Elections*
A_1	1st	1st	1st
A_2	2nd	2nd	2nd
A_3	3rd	3rd	3rd

	(b) Intransitive Outcome		
Alternative	*2013 Income*	*2015 Income*	*2016 Elections*
A_1	1st	2nd	3rd
A_2	2nd	3rd	1st
A_3	3rd	1st	2nd

that yields the greatest value. Transitive and intransitive choice situations are illustrated in Table 5.1.

Forms of Rationality

A rational choice is often defined as a choice that maximizes individual utility. One problem with this view, as defined by rational choice theorists,[7] is that it is not grounded in an empirically testable theory. A testable theory is one with hypotheses that can be compared with observations. Rational choice theory tends to remove choice from the process by simply assuming that persons maximize their utility according to the model. There is little or no room for choice, deliberation, or error. As such, it is an unrealistic model for analyzing policy choice.

Another theory is that of reasoned choice. People choose alternatives because they have reasons for choosing them. Rather than assuming that people make choices that maximize their utility, it is more useful to think of choices that maximize their utility for different reasons. Reasoned choices can be divided into five types: economic, technical, political, legal, and substantive. This reflects the fact that there are different values, norms, and institutional rules associated with identifiable modes of reasoning. In this context, any situation of choice can yield a prescribed course of action that is preferred to all others because it will result in desired outcomes. Yet most choice situations involve multiple stakeholders, uncertainty, and consequences that change over time. In fact, conflict and disagreement are essential characteristics of most policy choices. Attribute rankings conflict because they

[7]Jeffrey Friedman, ed. *The Rational Choice Controversy: Economic Models of Politics Reconsidered* (New Haven, CT: Yale University Press, 1996).

lie at the core of choices that must be made by public policymakers, mainly because these choices affect many individuals. Policy A might be better for one group, but policy B might be better for another. If time is relevant, policy B will be better in 10 or 20 years, but policy A might be superior as a better hedge against disaster.[8]

Perhaps one of the best descriptions of the current situation is *the scandal of rationality*. Regrettably, it may appear that the process of making prescriptions is not and cannot be "rational." Tempting as this conclusion might be, our inability to satisfy the conditions of the simple model of choice does not mean that the process of prescription is not and cannot be rational, if by *rationality* we mean a reflective process of reasoning. Contrary to the one-perspective theory of rationality, most choices have multiple attributes of rationality—they are multirational, and that is what makes them difficult and practically relevant. This means that simultaneous processes of reasoning underlie policy choices.[9]

1. *Technical rationality.* Technical rationality refers to reasoned choices based on a comparison of alternatives according to their *effectiveness.* The choice of solar and nuclear energy technologies based on a comparison of the number of BTUs of energy they produce is an example of technical rationality.
2. *Economic rationality.* Economic rationality refers to reasoned choices based on a comparison of alternatives according to their *efficiency.* Choices among alternative health care systems in terms of the difference between their total costs less total benefits (net benefits), or the criterion of net efficiency improvement, exemplify economic rationality.
3. *Legal rationality.* Legal rationality refers to reasoned choices involving a comparison of alternatives according to their *conformity to existing laws and statutes*. Choices involving the award of public contracts according to whether companies comply with laws against racial and sexual discrimination are an example of legal rationality.
4. *Social rationality.* Social rationality is a characteristic of reasoned choices based on a comparison of alternatives according to their maintenance of valued social institutions, that is, the extent to which they maintain or promote *institutionalization.* Choices involving the extension of rights to democratic participation in elections are an example of social rationality, because a democratic institution is promoted as a value.
5. *Substantive rationality.* Substantive rationality is a characteristic of reasoned choices that involve the comparison of *multiple forms of rationality*—technical, economic, legal, social—in order to make the most appropriate choice under given circumstances. Many issues of government information policy involve questions about the usefulness of new computer technologies, their costs and benefits to society, their legal implications for rights to privacy, and their consistency with democratic institutions. Debates about these issues may be characterized in terms of substantive rationality, which almost always involves tradeoffs.

---

[8]Friedman, p. 30.

[9]See Stephen Toulmin. *Return to Reason* (Cambridge, MA: Harvard University Press, 2009); Paul Diesing, *Reason and Society* (Urbana: University of Illinois Press, 1962); and Max Weber, *The Theory of Social and Economic Organization* (New York: Simon and Schuster, 1915/2009).

CRITERIA FOR POLICY PRESCRIPTION

The several types of rationality are associated with specific decision criteria for making prescriptions. By *decision criteria*, we mean explicit rules, principles, or standards used to justify policy choices. Decision criteria are of six main types: effectiveness, efficiency, adequacy, equity, responsiveness, and appropriateness.[10]

Effectiveness refers to the achievement of a valued outcome. Effectiveness, a property of technical rationality, is measured in terms of units of products or services or their monetary value. If more energy, defined in terms of British Thermal Units (BTUs), is produced by nuclear generators than solar collection devices, then nuclear generators are more effective. In this context, an *effective* health policy is one that provides people with more quality adjusted life years (QALYs).

Efficiency refers to the amount of effort required to produce a given level of effectiveness. Efficiency, which is a property of economic rationality, is calculated as the costs of producing a unit of product or service, for example, dollars per gallon of irrigation water or dollars per medical examination. Efficiency may also be calculated as the total monetary benefits per gallon of irrigation water less the total monetary costs per gallon of irrigation water. Another way to determine efficiency is to compare the *opportunity costs* of a policy with its rival. The policy that achieves the larger net benefits is said to be efficient.

Adequacy refers to a policy that achieves a defined threshold of effectiveness or efficiency. The criterion of adequacy specifies expectations about the strength of a relationship between a policy and a fixed level of effectiveness or efficiency. The criterion of adequacy and its relation to effectiveness and efficiency is illustrated by the four types of problems displayed in Table 5.2.

1. *Type I problems.* Problems of this type involve fixed costs and variable effectiveness. When maximum allowable budgetary expenditures specify fixed costs, the aim is to maximize effectiveness within the limits of available resources; for example, given a fixed budget of $1 million for each of two programs, a health policy that results in the greatest improvement in the QALYs in a community. The procedure for solving Type I problems is called *equal-cost analysis,* because the costs of alternatives that vary in effectiveness are equal. Here the most adequate policy is one that maximizes the achievement of outcomes while remaining within fixed costs.
2. *Type II problems.* Problems of this type involve variable costs and fixed effectiveness. When the level of outcomes is fixed, the aim is to minimize costs. For example, if public transportation facilities must serve at least 100,000 persons, the problem is to identify those alternatives—bus, monorail, subways—that will achieve this fixed level of effectiveness at least cost. The procedure for solving Type II problems is called *equal-effectiveness analysis,* because alternatives that are equally effective vary in costs. Here an adequate policy is one that minimizes costs while achieving a fixed level of effectiveness.

[10] On decision criteria, see Theodore H. Poister, *Public Program Analysis: Applied Research Methodology* (Baltimore, MD: University Park Press, 1978), pp. 9–15.

TABLE 5.2

Criteria of Adequacy: Four Types of Problems

	Costs	
Effectiveness	*Equal*	*Variable*
Variable	Type I Equal Cost- Variable Effectiveness	Type III Variable Cost- Variable Effectiveness
Equal	Type IV Equal Cost- Equal Effectiveness	Type II Variable Cost- Equal Effectiveness

3. *Type III problems.* Problems of this type involve variable costs and variable effectiveness. For example, the choice of an optimal budget to maximize the attainment of outcomes is a Type III problem. The procedure to solve Type III problems is called *variable-cost-variable-effectiveness analysis,* because costs and effectiveness can take any value. Here the most adequate policy is one that maximizes the ratio of effectiveness to costs.
4. *Type IV problems.* Problems of this type involve equal costs and equal effectiveness. Type IV problems are especially difficult to resolve, because analysis is not only limited by the constraint that costs not exceed a certain level; analysis is also limited by the constraint that there is a predetermined level of effectiveness. For example, if public transportation facilities must serve a minimum of 100,000 persons annually—yet costs have been fixed at an unrealistic level—then any policy alternative must either satisfy both constraints or be rejected. Many public contracts are Type IV problems.

The different definitions of adequacy contained in these four types of problems point to the complexity of relationships between costs and effectiveness. For example, two programs designed to provide municipal services—for example, recreation, waste disposal, police and fire protection—may differ in terms of both effectiveness and costs (Figure 5.1). Program I achieves a higher overall level of effectiveness than Program II, but Program II is less costly at lower levels of effectiveness. Should the analyst prescribe the program that maximizes effectiveness (Program I), or the program that minimizes costs at the same level of effectiveness (Program II)?

To answer this question, we must look at the relation between costs and effectiveness, rather than view costs and effectiveness separately. However, this is where complications begin. (1) If we are dealing with a Type I (equal-cost-equal-effectiveness) problem and costs are fixed at \$20,000 (C_2), Program II is more adequate because it achieves the highest level of effectiveness while remaining within the fixed-cost constraint. (2) If we are faced with a Type II (equal-effectiveness-variable-cost) problem and effectiveness is fixed at 6,000 units of service (E_2), Program I is more adequate. (3) If, on the other hand, we are dealing with a Type III (variable-cost-variable-effectiveness) problem, where costs and effectiveness are

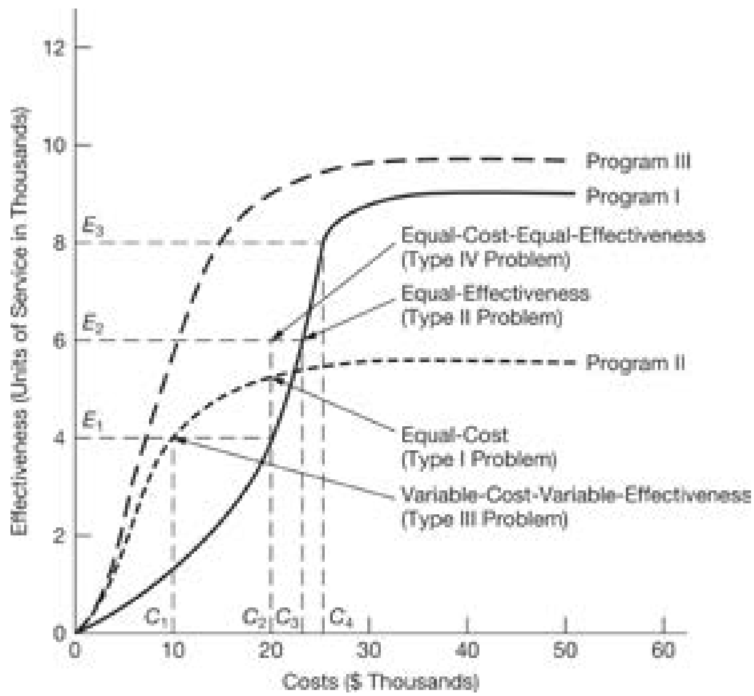


FIGURE 5.1
Cost–effectiveness comparisons using four criteria of adequacy

Source: Adapted from Quade (1975): 93.

free to vary, Program II is more adequate, because the ratio of effectiveness to costs (*effectiveness–cost ratio*) is greatest at the intersection of E_1 and C_1. Here Program II produces 4,000 units of service for $10,000, that is, a ratio of 4,000 to 10,000 or 0.4. By contrast, Program I has an effectiveness–cost ratio of 0.32 (8,000 units of service divided by $25,000 = 8,000/25,000 = 0.32).[11] Finally, (4) if we are dealing with a Type IV (equal-cost-equal-effectiveness) problem, where both effectiveness and costs are fixed at E_2 and C_2, neither program is adequate. This dilemma, which permits no adequate solution, is known as *criterion overspecification.*

The lesson of this illustration is that it is not possible to choose between alternatives on the basis of *either* costs *or* effectiveness. It is the specific relation between costs and effectiveness that matters. While it is often possible to convert measures of effectiveness into monetary benefits, which permits us to calculate net income benefits by

[11] A simple way to display effectiveness–cost ratios graphically is to draw a straight line from the origin (i.e., where *x* and *y* axes intersect) of a graph to the knee of the effectiveness–cost curve (i.e., the point where the straight line drawn from the origin touches the curvature but does not pass through it). The more that the line moves leftward towards the vertical effectiveness scale, the greater the ratio between effectiveness and costs.

subtracting monetary costs from monetary benefits, it is difficult to establish convincing dollar equivalents for many of the most important policy outcomes. What is the dollar equivalent of a life saved through traffic safety programs? What is the dollar value of international peace and security promoted by United Nations educational, scientific, and cultural activities? What is the dollar value of natural beauty preserved through environmental protection legislation? While such questions will be examined further when we discuss cost–benefit analysis, it is important to recognize that the measurement of effectiveness in dollar terms can be complex and difficult.[12]

Sometimes it is possible to identify an alternative that simultaneously satisfies all criteria of adequacy. For example, the upper broken-line curve in Figure 5.1 (Program III) represents a third program that adequately meets equal cost as well as equal effectiveness criteria and also has the highest ratio of effectiveness to costs. As this situation is rare, it is almost always necessary to specify the level of effectiveness and costs that are regarded as adequate. Yet this is a matter of judgment about what constitutes an adequate level of effectiveness, given the costs.

Questions of adequacy cannot be resolved by arbitrarily adopting a single criterion. For example, net income benefits (dollars of effectiveness minus dollar costs) are *not* an appropriate criterion when costs are fixed and a single program with the highest benefit–cost ratio can be repeated many times within total fixed-cost limits. This is illustrated in Table 5.3, where Program I can be repeated 10 times up to a fixed-cost limit of $40,000, with total net benefits of $360,000 ($36,000 × 10). Program I, which can be repeated 10 times, has the highest benefit–cost ratio. But if Program I cannot be repeated—that is, if only one of the three programs must be selected—Program III should be chosen. Program III produces the greatest net benefits, although it has the lowest benefit–cost ratio.

Returning to the five types of rationality discussed earlier, the criterion of *equity* is closely related to legal and social rationality. In its most basic form, equity refers to the distribution of outcomes and effort among different groups in society. An equitable policy is one where outcomes (e.g., units of service) or efforts (e.g., investment) are fairly or equitably distributed. Other types of distribution involve income or wealth. In the United States, the percentage share of household income of the top 20 percent of households increased from 43 to 51 percent between 1966 and 2012, while the share of all other households fell from roughly 57 to 49 percent in the same period.[13] Inequality of household income, measured by the Gini coefficient, increased by 20 percent, from .40 to .48 in the same period. In 2010, the Gini coefficient (after taxes and income transfers) for the United States was the highest among ten western countries. These measures show that, apart from efficiency, the United States has the most inequitable distribution of income in the western world.[14] This is consistent with Thomas Piketty's comparisons of income equality in Europe and the United States between 1910 and 2010 (Figure 5.2).[15]

[12]See, for example, Richard Zeckhauser, "Procedures for Valuing Lives," *Public Policy* (Fall 1975): 419–464, and Baruch Fischhoff, "Cost-Benefit Analysis and the Art of Motorcycle Maintenance," *Policy Sciences* 8 (1977): 177–202.

[13]Russell Sage Foundation, *Chartbook of Social Inequality*. New York: Russell Sage Foundation, 2013.

[14]Thomas Piketty, *Capital in the 21st Century* (Cambridge, MA: Harvard University Press, 2014).

[15]Piketty's data is in the technical appendix available at http://piketty.pse.ens.fr/capital21c.

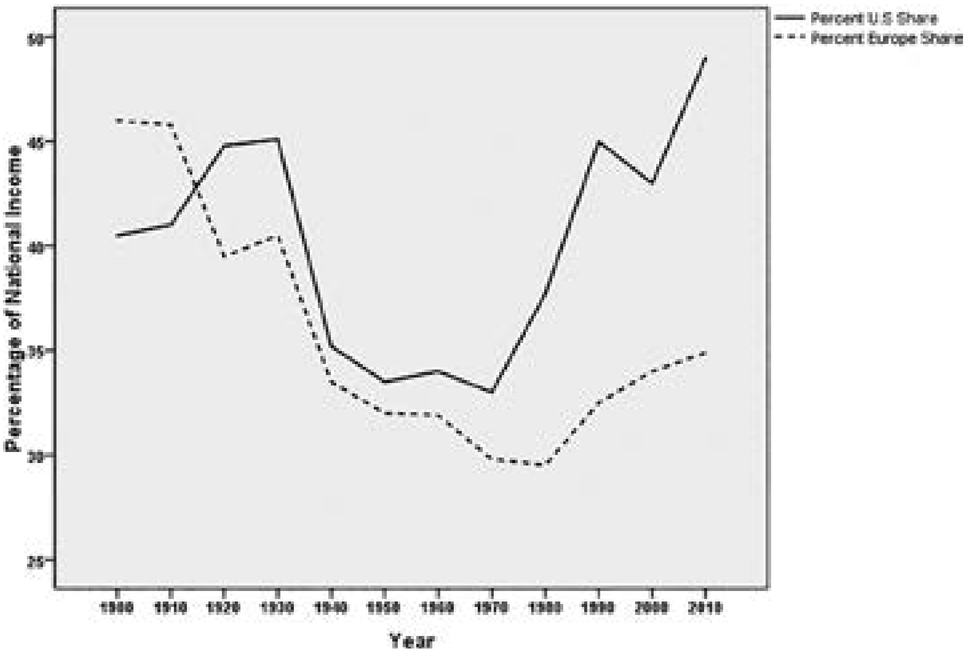


FIGURE 5.2
Income Inequality: Percentage Share of Top 10 Percent of U.S. and European Households in National Income, 1900–2010

Source: Piketty (2014), technical appendices.

Policies designed to redistribute income, educational opportunity, or public services are prescribed on the basis of the criterion of equity. A given program might be effective, efficient, and adequate—for example, the benefit–cost ratio and net benefits may be superior to all other programs—yet it might still be rejected on grounds that it will produce an inequitable distribution of costs and benefits. This could happen under several conditions: Those most in need do not receive services in proportion to their numbers, those who are least able to pay bear a disproportionate share of costs, or those who receive most of the benefits do not pay the costs.

The criterion of equity is closely related to competing conceptions of justice or fairness and to ethical issues surrounding the appropriate basis for distributing resources in society. Such problems of "distributive justice," which have been widely discussed since the time of Aristotle and Kautilya, may occur when a course of action that affects two or more persons in society is prescribed. In explicitly defining objectives for society as a whole, the analyst may actually be seeking a way to measure *social welfare,* that is, the aggregate satisfaction experienced by members of a community.

However, individuals and groups have different values. What satisfies one person or group often does not satisfy another. Under these circumstances, the analyst must answer a fundamental question: How can a policy maximize the welfare of society, and not just the

TABLE 5.3

Comparison of Benefit–Cost Ratio and Net Benefits as Criteria of Adequacy (in Thousands of Dollars)

Program	Benefits	Costs	Benefit–Cost Ratio	Net Benefits
I	40	4	10	36
II	64	8	8	56
III	100	40	2.5	60

Source: Adapted from Zeckhauser and Schaefer (1968): 73.

welfare of particular individuals or groups? The answer to this question may be pursued in several different ways:

1. *Maximize individual welfare.* Here we attempt to maximize the welfare of all individuals simultaneously. This requires that a single transitive preference ranking be constructed on the basis of all individual values. *Arrow's impossibility theorem,* as we have seen, demonstrates that this is impossible even in cases where there are two persons and three alternatives.
2. *Protect minimum welfare.* Here the analyst can attempt to increase the welfare of some persons while still protecting the positions of persons who are worst off. This approach is based on the *Pareto criterion,* which states that one social state is better than another if at least one person is better off, and no one is worse off. A *Pareto optimum* is a social state in which it is not possible to make any person better off without also making another person worse off. The Pareto criterion is very rarely applicable, because most policy decisions involve the provision of services to persons who are made better off by taxing those who are made worse off.
3. *Maximize net welfare.* Here we attempt to increase net welfare (total benefits less total costs) but assume that the gains could be used to compensate losers. This approach is based on the *Kaldor–Hicks criterion:* One social state is better than another if there is a net gain in efficiency (total benefits minus total costs) and if those who gain can compensate losers. For all practical purposes, this criterion, which does not require that losers actually be compensated, avoids the issue of equity. The Kaldor–Hicks criterion is one of the foundations of policies that attempt to provide a social safety net.
4. *Maximize redistributive welfare.* Here we attempt to maximize redistributional benefits to selected groups in society, for example, the racially oppressed, poor, or sick. One redistributive criterion has been put forth by philosopher John Rawls: one social state is better than another if it results in a gain in welfare for members of society who are worst off.[16]

Rawls's formulation attempts to provide an ethical foundation for the concept of justice. It does so by requesting that we imagine ourselves in an "original" state, where there is a "veil of ignorance" about the future distribution of positions, statuses, and resources in a future

[16]John Rawls, *A Theory of Justice* (Cambridge, MA: Harvard University Press, 1971).

civil society. In this "original" state, individuals are assumed to choose a social order based on the redistributive criterion, because it is in everyone's individual interest to establish a society in which they will not be worst off.

By postulating this "original" condition, it becomes possible to reach consensus on a just social order. This "original" condition should be contrasted with the conditions of present societies, where vested interests make it impossible to reach consensus on the meaning of justice. The weakness of Rawls's formulation is its oversimplification or avoidance of conflict. The redistributive criterion is appropriate for well-defined problems and not for the types of problems typically encountered in public policy. While this does not mean that the redistributive criterion cannot be used to make choices, it does mean that we have still not reached a single basis for defining social welfare.

Hence, none of these criteria of equity is fully satisfactory. The reason is that conflicting views about the rationality of society and social institutions (social rationality) or of the appropriateness of legal norms guaranteeing rights to property (legal rationality) cannot be resolved simply by appealing to formal economic criteria (e.g., the Pareto or Kaldor–Hicks criteria) or to formal philosophical principles (e.g., Rawls's redistributive criterion). Questions of equity, fairness, and justice are political, that is, they involve the distribution and legitimation of power in society. While economic theory and moral philosophy can improve our capacity to critically assess competing concepts of equity, they cannot replace the political process.

Responsiveness refers to the extent that a policy satisfies the needs, preferences, or interests of particular groups. The criterion of responsiveness is important because an analyst can satisfy all other criteria—effectiveness, efficiency, adequacy, equity—yet still fail to respond to the actual needs of a group that is supposed to benefit from a policy. A recreation program might result in an equitable distribution of athletic activities but be unresponsive to the needs of particular groups (e.g., the elderly). In effect, the responsiveness criterion asks a practical question: Do criteria of effectiveness, efficiency, adequacy, and equity actually reflect the perceived needs, preferences, and values of particular groups?

The final criterion is that of *appropriateness*. This criterion is intimately related to substantive rationality, because questions about the appropriateness of a policy are not concerned with individual criteria but two or more criteria taken together. Appropriateness refers to the value or worth of a program's outcomes and to the tenability of assumptions underlying these objectives. While all other criteria take objectives for granted—for example, the value either of efficiency or of equity—the criterion of appropriateness asks whether these objectives are proper ones for society. To answer this question, all criteria must be considered together—that is, there must be an inquiry into multiple forms of rationality—and consider the application of higher-order criteria (meta-criteria) that may justify those of effectiveness, efficiency, adequacy, equity, and responsiveness.[17]

The criterion of appropriateness is necessarily open ended, a matter of inquiry rather than testing, because by definition it is intended to go beyond any set of existing criteria.

[17]On meta-criteria, including generality, consistency, and clarity, see Duncan MacRae Jr., *The Social Function of Social Science* (New Haven, CT: Yale University Press, 1976).

For this reason, there is not and cannot be a standard definition of appropriateness. However, we can consider several examples:

1. *Equity and efficiency.* Is equity as redistributional welfare (Rawls) an appropriate criterion when programs designed to redistribute income to the poor are so inefficient that only a small portion of redistributional benefits actually reaches them? When such programs are viewed as a "leaky bucket" used to carry benefits to the poor, we may question whether equity is an appropriate criterion.[18]
2. *Equity and entitlement.* Is equity as minimum welfare an appropriate criterion when those who receive additional benefits have not earned them through socially legitimate means? Analysts may question the appropriateness of the Pareto criterion when those who gain (although none lose) have done so through corruption, fraud, discrimination, and unearned inheritance.[19]
3. *Efficiency, equity, and humanistic values.* Are efficiency and equity appropriate criteria when the means required to achieve an efficient or just society conflict with democratic processes? Analysts may challenge the appropriateness of efficiency or equity when efforts to rationalize decision-making subvert conditions required for emancipation, individuation, or self-actualization.[20] Efficiency, equity, and humanism are not necessarily equivalent. Alienation, as Marx and other social thinkers have recognized, is not automatically eliminated by creating an abundant society of equals.
4. *Equity and reasoned ethical discourse.* Is the conception of equity as redistributive welfare (Rawls) appropriate when it subverts opportunities for reasoned ethical discourse? This notion of equity as redistributive welfare can be challenged on grounds that it presupposes an individualistic conception of human nature where ethical claims are no longer derived from reasoned arguments, but are "the contractual composite of arbitrary (even if comprehensible) values individually held and either biologically or socially shaped. The structure of the Rawlsian argument thus corresponds closely to that of instrumental rationality; ends are exogenous, and the exclusive office of thought in the world is to ensure their maximum realization. [Rawls's premises] reduce all thought to the combined operations of formal reason and instrumental prudence in the service of desire."[21]

APPROACHES TO PRESCRIPTION

In making policy prescriptions, a number of interrelated questions must be addressed. Whose needs, values, and opportunities are at issue, and what alternatives are available

[18]See Arthur M. Okun, *Equality and Efficiency* (Washington, DC: Brookings Institution, 1975).
[19]See, for example, Peter Brown, "Ethics and Policy Research," *Policy Analysis* 2 (1976): 325–340.
[20]See, for example, Jurgen Habermas, *Toward a Rational Society* (Boston, MA: Beacon Books, 1970), and *Legitimation Crisis* (Boston, MA: Beacon Books, 1975).
[21]See Laurence H. Tribe, "Ways Not to Think about Plastic Trees," in Tribe and others, *When Values Conflict* (Cambridge, MA: Ballinger, 1976), p. 77.

for their satisfaction? What goals and objectives should be attained, and how should they be measured? How much will it cost to attain objectives, and what kinds of constraints—budgetary, legal, administrative, political—may impede their attainment? Are there side effects, spillovers, and other anticipated and unanticipated consequences that should be counted as costs or benefits? How will the value of costs and benefits change over time? How certain is it that forecasted outcomes will occur? What should be done?

Public versus Private Choice

Answers to these questions are frequently just as relevant for policymaking in the private as in the public sector. However, there are several important differences between these sectors[22]:

1. *Nature of public policy processes.* Policymaking in the public sector involves bargaining, compromise, and conflict among citizens' groups, legislative bodies, executive departments, regulatory commissions, businesses, and other stakeholders. There is no single producer or consumer of goods and services whose profits are to be maximized. The presence of stakeholders with competing values makes problems of choice more complex in the public than in the private sector.
2. *Collective nature of public policy goals.* Policy goals in the public sector are collective, which means that they are supposed to reflect society's interests, or some broad conception of the "public interest." The specification of these collective goals, as we have seen, often involves multiple competing criteria, from effectiveness, efficiency, and adequacy to equity, responsiveness, and appropriateness.
3. *Nature of public goods.* Public and private goods may be divided into three groups: specific goods, collective goods, and quasi-collective goods. *Specific goods* are exclusive, because the person who owns them has the legal right to exclude others from their benefits. The allocation of specific goods (for example, cars, merchandise, or private housing) is made on the basis of market prices, which are determined by supply and demand. *Collective goods* are nonexclusive, because they may be consumed by everyone. No person can be excluded from the consumption of air, water, and roads provided by the government. It is usually not impossible to allocate collective goods on the basis of market prices, because relationships between supply and demand do not operate the same way in the public sector. *Quasi-collective goods* are specific goods whose production has significant spillover effects for society. Although education can be provided by the private sector, its spillover into areas such as employment and national defense is regarded as so important that governments produce large quantities at a cost affordable by all.

[22]H. H. Hinrichs and G. M. Taylor, *Systematic Analysis: A Primer on Benefit-Cost Analysis and Program Evaluation* (Pacific Palisades, CA: Goodyear Publishing, 1972), pp. 4–5; Dennis C. Mueller, "Public Choice: An Introduction," in *Readings in Public Choice and Constitutional Political Economy* (New York: Springer, 2008): pp. 31–46.

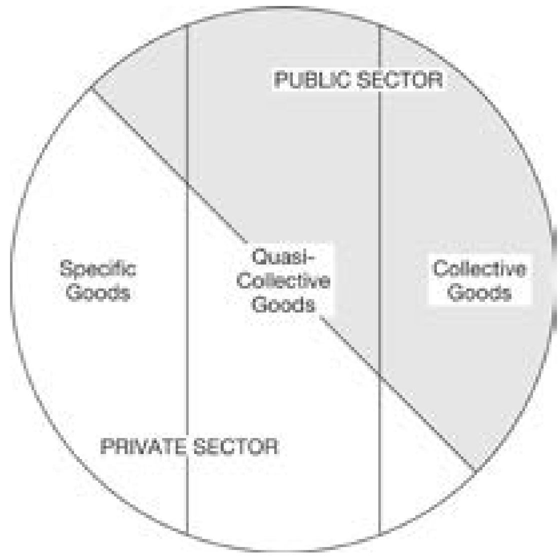


FIGURE 5.3
Three Types of Goods in the Public and Private Sectors

Source: Hinrichs and Taylor (1972): 5.

Organizations in the public and private sectors produce each of these three types of goods. Nevertheless, the public sector is primarily committed to the provision of such collective and quasi-collective goods as education, social welfare, public safety, transportation, environmental protection, and energy conservation. By contrast, the private sector is chiefly concerned with the production of such specific goods as food commodities, appliances, machinery, and housing. Contrasts between the production of goods in the public and private sectors are illustrated in Figure 5.3.

Because the nature of these three types of goods differs, so do the criteria for estimating their value. The primary criterion of a private firm producing specific goods for the market is to make profits, that is, to maximize the difference between the total revenues earned by selling a product and the total costs required for its production. When a private firm is faced with a choice between two or more products that differ in the revenues they will earn and the costs required for their production, the firm will choose that product that maximizes profits, defined as total revenue minus total costs. If the firm should decide to invest in a product that yields lower profits, the opportunity cost of the decision may be calculated. An *opportunity cost* is the benefits forgone by investing resources in one product or service when another more profitable alternative might have been chosen.

Supply and Demand

Opportunity costs in the private sector can be calculated by using market prices as a measure of costs and benefits. Market prices of specific goods are determined by supply and demand. If we look at various combinations of the price and quantity of a specific good, we observe that consumers will demand greater quantities (Q) of a product as the price (P) for that product decreases and producers will supply greater quantities (Q) of a product as the price (P) of that product increases (Figure 5.4). Finally, the combination of price and quantity that yields a single level of consumer demand and producer supply—that is, the point where supply and demand intersect (P_eQ_e in Figure 5.4)—indicates the price and quantity of specific goods that will be sold on the market. Graphic representations of the various combinations of price and quantity where consumers and producers are willing to buy and sell a product are called *demand curves* and *supply curves.* The point where demand and supply curves intersect is called the *equilibrium price–quantity combination.*

The equilibrium price–quantity combination represents that point where consumers and producers, if they are completely free to choose what they will buy and sell, will produce and consume equal quantities of a good at a given price. If we know the equilibrium price–quantity combination for a specific good (e.g., electric toothbrushes), we can determine the profit of a given investment by subtracting the total costs required to produce a specific good from the total revenues earned by selling the product at the equilibrium price (P_e) and equilibrium quantity (Q_e). By knowing the profits that can be earned from alternative investments, it is possible to estimate the opportunity costs of investing resources to produce one product when another more profitable investment might be made. For example, if a firm producing electric toothbrushes (a specific good) and earning an annual profit of $1.2 million finds that it could have earned a profit of $1.5 million by producing electric drills at the same cost (including costs for additional labor, technology, or marketing) the opportunity cost of the investment in electric toothbrush production is $300,000.

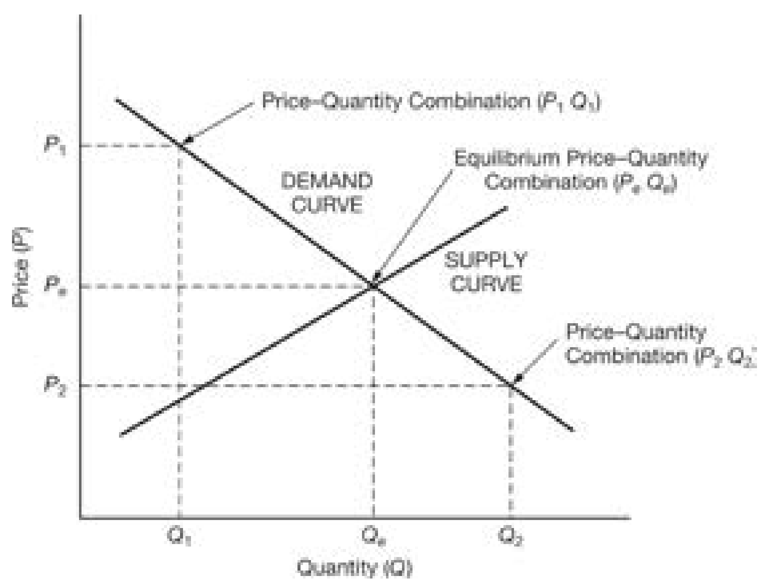


FIGURE 5.4
Supply and Demand Curves and the Equilibrium Price–Quantity Combination

The logic of profit maximization in the private sector can be extended to public policy. We can view public agencies as if they were private firms attempting to maximize profits on investments. Instead of using profits (total revenue minus total cost) as a criterion, net benefits (total benefits minus total costs) are used. We can also apply the concept of opportunity costs by viewing public programs as investments in the production of goods that might have been made by private firms. For example, if government invests $50 million in the construction of a dam that will yield $8 million in net benefits to farmers and other beneficiaries, it must extract these investment resources from private citizens who could have invested the $50 million elsewhere. If private investment would have yielded $9.5 million in net benefits (profits) to private citizens, the opportunity cost of building the dam is $1.5 million (net private benefits of $9.5 million minus net public benefits of $8 million).

Public Choice

The logic of profit maximization in the private sector begins to break down when we consider differences between public and private choice, including contrasts among specific, quasi-collective, and collective goods. While the logic of profit maximization might be applied to certain kinds of public goods (e.g., the production of hydroelectric power), there are reasons why concepts of profit and opportunity costs are difficult to apply to problems of public choice:

1. *Multiple legitimate stakeholders.* Public policymaking involves multiple stakeholders whose claims on public investments are guaranteed by law. While there are also stakeholders in private policymaking, only owners and stockholders can legitimately make claims. In the public sector, with its many legitimate stakeholders, it is difficult to know whose benefits should be maximized and who should bear the costs of public investments, typically through taxation.
2. *Collective and quasi-collective goods.* Because most public goods are collective (e.g., clean air) or quasi-collective (e.g., education), it is difficult or impossible to sell them on the market, where transactions are made on the basis of ability to pay the market price. However, market prices are frequently unavailable as a measure of net benefits or opportunity costs. Even where prices are based on surveys of what citizens are willing to pay, some persons regularly indicate that they are unwilling to pay for a public good but later use the good at a price lower than what they are actually willing to pay. This is called the *free-rider problem.* The free-rider problem also applies to businesses that pay no taxes but use the roads, sewers, and bridges paid for by taxpayers.
3. *Limited comparability of income measures.* Even when the net benefits of public and private investments can be expressed in a common unit of measure such as dollars, private investments with higher net benefits are not always preferable to public benefits with lower (or even zero or negative) net benefits. For example, a private investment in a new office building that yields $1 million in net benefits is unlikely to be preferred over a public investment in a cure for cancer that yields zero or negative net benefits. Even where comparisons are confined to public investments, the use of

income measures to prescribe policies implies that the goal of increasing aggregate income is more important than good health or better education, if these cannot be produced at an equal or lower cost. These other collective and quasi-collective goals are wrongly viewed as illegitimate simply because they fail to increase future income.[23]

4. *Social responsibility for social costs and benefits.* Private firms are responsible for their own *private costs* and *private benefits* but not for social ones, except as prescribed by law (e.g., antipollution legislation). By contrast, the costs and benefits of public programs are socialized, not privatized, thus becoming *social costs* and *social benefits* that go beyond the need to satisfy private owners and stockholders. Social costs and social benefits (e.g., the costs of damaging the environment by building highways or the benefits of enjoying the beauty of national parks) are difficult to quantify and often have no consensually defined market price. Such social costs and social benefits are called *intangibles* that cannot be satisfactorily measured.

Contrasts between public and private choice do not mean that the logic of profit maximization is entirely inapplicable to public problems. Contrasts do mean, however, that there are limits to the logic of privatization when applied to public problems. These limits as well as strengths are evident when we consider two of the most important approaches to prescription: cost–benefit analysis and cost–effectiveness analysis.

Cost–Benefit Analysis

Cost–benefit analysis permits comparisons of policies by quantifying their total monetary costs and total monetary benefits. When cost–benefit analysis is used to prescribe preferred policies, it is applied prospectively (*ex ante*). Cost–benefit analysis is also used to evaluate policies after they have been implemented (*ex post*). Modern cost–benefit analysis is based on *welfare economics*, which attempts to maximize social welfare, or the aggregate satisfaction experienced by members of a community.[24] The field of *welfare economics* is concerned with the ways that public investments contribute to the maximization of net income as a measure of the aggregate satisfaction (welfare) of society.

Cost–benefit analysis (hereafter abbreviated as CBA) has been applied to many kinds of public policies and programs. The idea of an economic accounting scheme incorporating costs and benefits appears to have originated with Albert Gallatin, U.S. Secretary of the Treasury who was active in developing natural resources in the western United States; Jules Dupuit, a French engineer; and the economist Alfred Marshall, who formulated most of the formal concepts that are the foundation of cost–benefit analysis and modern utility theory. However, the practical development of CBA was a consequence of the Federal Navigation

[23]Alice Rivlin, *Systematic Thinking for Social Action* (Washington, DC: Brookings Institution, 1971), p. 56.

[24]Comprehensive treatments are provided by Edward M. Gramlich, *Benefit-Cost Analysis of Public Programs*, 2nd ed. (Englewood Cliffs, NJ: Prentice Hall, 1990); Anthony E. Boardman, David H. Greenberg, Aidan R. Vining, and David L. Weimer. *Cost-Benefit Analysis: Concepts and Practice*, 4th ed. (Upper Saddle River, NJ: Prentice Hall, 2011); and Mark A. Moore, David L. Weimer, Mark A. Moore, and D. H. Greenberg, "'Just Give Me a Number!' Practical Values for the Social Discount Rate," *Journal of Policy Analysis and Management* 23, 4 (2004): 789–812.

Act of 1936, which directed that the U.S. Army Corps of Engineers carry out projects for the improvement of waterways only when the total benefits exceeded the costs. The earliest applications of cost–benefit analysis were in the area of hydroelectric dam construction and the provision of water resources, including efforts to analyze the costs and benefits of hydroelectric power, flood control, irrigation, and recreation. Other more recent applications include transportation, health, manpower training, and urban renewal.

When used to make prescriptions in the public sector, CBA has several characteristics:

1. Cost–benefit analysis seeks to measure *all* costs and benefits to society that may result from a public policy or program, especially intangibles that cannot be easily measured according to monetary costs and benefits.
2. Traditional CBA epitomizes *economic rationality,* because the criterion most frequently employed is global economic efficiency, which refers to the ratio of benefits to costs. A policy is said to be efficient if its *net benefits* (i.e., total benefits minus total costs) are greater than zero (B > C and B/C > 0) and higher than those that would have resulted from an alternative public or private investment.
3. Traditional CBA uses the *private marketplace* as a point of departure in prescribing policies. The opportunity costs of a public investment are calculated on the basis of what net benefits might have been gained by investing in the private sector.
4. Contemporary CBA, sometimes called *social cost–benefit analysis,* can be used to measure redistributional benefits. As social cost–benefit analysis is concerned with criteria of equity, it is consistent with *social rationality.*

Strengths of CBA. Cost–benefit analysis has a number of strengths. First, *both* costs and benefits are measured in dollars, euros, or some other established currency as a common unit of value. This permits analysts to subtract costs from benefits, a task that is not possible with cost–effectiveness analysis. Second, CBA permits us to go beyond the confines of a single policy or program and link benefits to enlarging the income of the economic system of society as a whole. This is possible because the results of individual policies and programs can be expressed in monetary terms.

Finally, CBA allows analysts to compare policies in widely differing areas (for example, health, transportation, education, child care), because net efficiency benefits (benefits minus costs) are expressed in terms of a common unit of currency. This is not possible when effectiveness is measured in terms of units of service, because the number of persons treated by physicians, for example, cannot be directly compared with the number of miles traveled on roads.

Limitations of CBA. Cost–benefit analysis, in both its traditional and its contemporary forms, also has several limitations. First, an exclusive emphasis on economic efficiency means that criteria of equity are meaningless or inapplicable. In practice, the Kaldor–Hicks criterion simply ignores problems of redistributional benefits, while the Pareto criterion seldom resolves conflicts between efficiency and equity. Second, monetary value is an inadequate measure of responsiveness, because the value of income varies from person to person. For example, an extra $100 of income is far more significant to the head of a poor household than to a millionaire. This problem of *limited interpersonal comparisons* means that income is an inappropriate measure of individual satisfaction and social welfare. Third, when market prices are unavailable for important goods (e.g., clean air or lives saved on highways), analysts are often forced to estimate *shadow prices,* that is, subjective estimates of the price that citizens might be willing to pay for goods and services. These subjective judgments may simply be arbitrary.

Cost–benefit analysis, even when it takes into account problems of redistribution and social equity, is closely tied to income as a measure of satisfaction. For this reason, it is difficult to discuss the appropriateness of *any* objective that cannot be expressed in monetary terms. This exclusive concentration on net income or redistributional benefits inhibits reasoned debates about the ethical or moral bases of alternative policies. CBA is therefore limited in its ability to consider relations among alternative forms of reason—technical, economic, social, legal—as part of an overall effort to establish the substantive rationality of policies. Hence, CBA may provide us with "the price of everything and the value of nothing."[25]

Types of Costs and Benefits

In using CBA, it is essential to consider *all* costs and *all* benefits that may result from a policy or program. While such a comprehensive inventory of costs and benefits is difficult to achieve in practice, it mitigates errors that occur when some costs and benefits are omitted from an analysis. One way to guard against such errors is to classify costs and benefits so that none will be missed: inside ("internal") versus outside ("external"), directly measurable ("tangible") versus indirectly measurable ("intangible"), primary ("direct") versus secondary ("indirect"), and net efficiency ("real") versus distributional ("pecuniary"). These types of costs and benefits, and the questions on which they are based, are illustrated in Figure 5.5.

Inside versus outside costs and benefits. Here the question is whether a given cost or benefit is internal or external to a given jurisdiction. Inside costs and benefits are called *internalities,* while outside or external ones are called *externalities.* What is an inside cost or benefit (internality) in one case will be an outside one (externality) in another. The difference depends on how the boundaries are drawn around jurisdiction, for example, a local government. If the boundary is society as a whole, there may be no externalities. If, however, the boundary is a particular target group or jurisdiction, there will be internalities as well as externalities. Externalities are the positive and negative spillovers outside the boundaries of the jurisdiction or target group. For example, the construction of high-rise apartments in a central-city area as part of an urban renewal program has certain costs and benefits within the urban jurisdiction, including expenditures on construction and income derived from rents. The same urban renewal program also has external costs to suburban jurisdictions that must provide additional fire and police services in areas where vagrants or criminals have resettled because they can no longer afford to live in the central-city area.

[25]This quotation is attributed to Oscar Wilde, the Irish classical writer, novelist, and social critic. An anonymous economist retorted: "But, Mr. Wilde, prices are values."

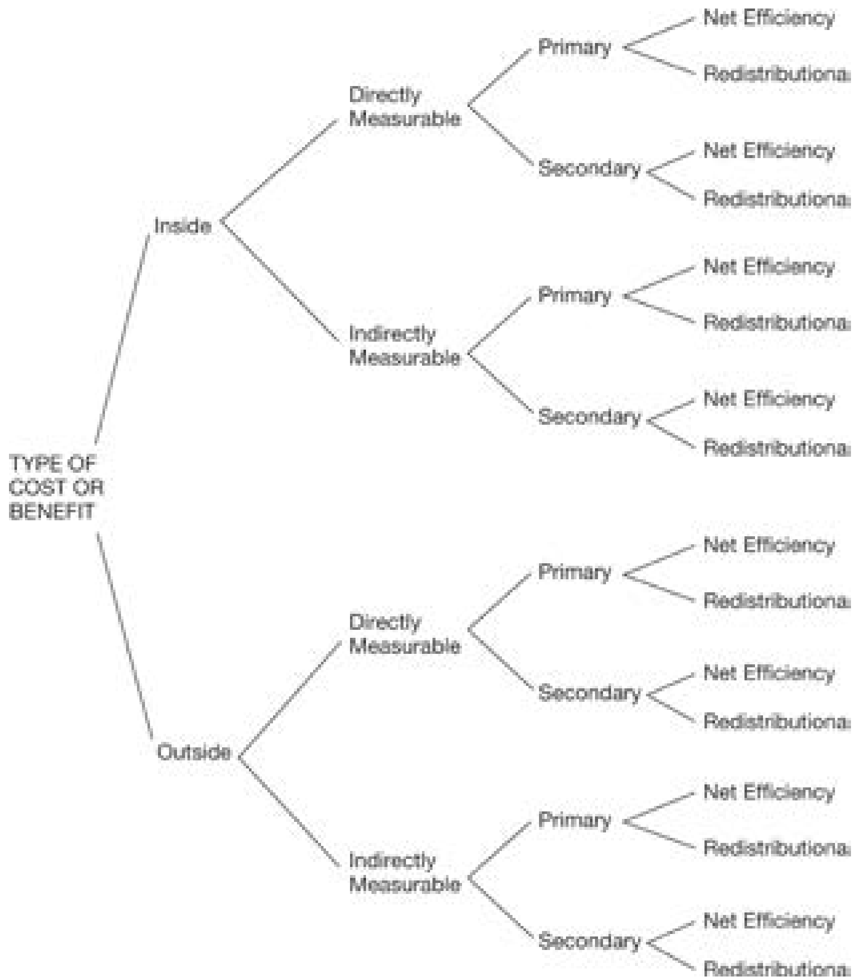


FIGURE 5.5
Classification of Costs and Benefits According to Four Types of Questions

Directly measurable versus indirectly measurable costs and benefits. The question here is whether the cost or benefit is a "tangible" or "intangible" one. *Tangibles* are costs and benefits that are directly measurable in terms of known market prices for goods and services, while *intangibles* are costs and benefits that are indirectly measurable in terms of estimates of such market prices. When dealing with intangibles, such as the price of clean air, the analyst may attempt to estimate shadow prices by making a subjective judgment about the dollar value of costs and benefits. In the urban renewal example, the analyst may try to estimate the opportunity costs that various groups might be willing to pay for the destruction of the sense of community resulting from urban renewal.

Primary and secondary costs and benefits. Here the question is whether the cost or benefit is a "direct" or "indirect" result of a policy. A *primary cost or benefit* is one that is related to the most highly valued program objectives, while a *secondary cost or benefit* is one that is related to objectives that are less valued. For example, an urban renewal program may have as its most important objective the provision of low-cost housing to the poor, in which case the primary costs and benefits would include expenditures for construction and income from rents. Secondary costs and benefits would involve intangible costs, such as the destruction of a sense of community, or tangible benefits, such as reduced costs of police and fire services because of better street lighting and fire-resistant construction.

Net efficiency versus redistributional benefits. Here the question is whether combined costs and benefits create an increase in aggregate income or result merely in shifts in income or other resources among different groups. *Net efficiency benefits* are those that represent a "real" increase in net income (total benefits minus total costs), while *redistributional benefits* are those that result in a "pecuniary" shift in the incomes of one group at the expense of another but without increasing net efficiency benefits. Such changes are called *real benefits* and *pecuniary benefits,* respectively. For example, an urban renewal project may produce $1 million in net efficiency benefits. If urban renewal also results in increased sales in small grocery stores in the immediate area—and decreased sales in stores that are more distant from new high-rise apartments—the benefits and the costs of income gained and lost are "pecuniary." They cancel each other out without producing any change in net efficiency benefits.

To call one type of benefit "real" and the other merely "pecuniary" can introduce considerable bias: Is an increase in the disposable income of the poor less "real" than an improvement in the net benefits of the community? For this reason, it may be best to drop the categories of "real" and "pecuniary" altogether, provided that we do not *add* any increase in redistributional benefits to net efficiency benefits without also *subtracting* the costs of redistribution to those whose income or other benefits were reduced. This error, called *double counting,* can produce invalid estimates of net efficiency benefits.

Answers to these four questions can result in many combinations of costs and benefits. A given inside cost or benefit may be directly measurable (tangible) or indirectly measurable (intangible). In turn, an intangible cost or benefit may be primary or secondary, depending on its relative importance. A primary or secondary objective may be defined in terms of net efficiency or redistribution. Typically, either net efficiency or redistribution must be a primary objective, because it is not always possible to increase net efficiency benefits *and* redistributional benefits at the same time. In other words, these two types of benefits may conflict. They represent a situation requiring *tradeoffs,* that is, conscious efforts to determine how much of one objective should be sacrificed to obtain another. However, this relation between real and pecuniary benefits is not an immutable law. Not only are some firms and industries able to increase both at the same time, but the history of this issue that has been documented in classics such as *The Great Transformation*[26] suggests that markets in some societies have been based on the value of redistribution, with the value of what we now call "real" efficiency benefits a necessary but secondary condition of redistribution.

[26]Karl Polanyi, *The Great Transformation: The Political and Economic Origins of Our Time* (Boston, MA: Beacon Press, 1944, 2001), pp. 54–58. The book is regarded by many as one of the most original and important books of the twentieth century.

Tasks in Cost–Benefit Analysis

In conducting a cost–benefit analysis (CBA), most of the following tasks (Table 5.5) are important for making maximally plausible prescriptions. One of the important tasks is converting goals, or general purposes, into specific objectives (Table 5.4).

Problem structuring. The search for alternative formulations of the problem and the definition of the boundaries of the metaproblem (Chapter 3) are essential tasks in CBA. Problem structuring does not occur once, but reoccurs throughout the analysis. Problem structuring yields information about potentially relevant goals, objectives, alternatives, criteria, target groups, costs, and benefits.

Specification of objectives. Analysts usually begin with general aims or goals, for example, controlling cocaine addiction. Goals, must be converted into objectives that are temporally specified and measurable (see Table 5.5). The goal of controlling cocaine addiction may be converted into a number of specific objectives, for example, a 50 percent reduction in the supply of cocaine within 5 years. Objectives usually imply policy alternatives. The reduction in supply implies a policy of drug interdiction, whereas another objective—for example, a 50 percent reduction in the demand for cocaine—not only implies a policy of drug interdiction to raise the price of cocaine and lower demand, but also suggests drug rehabilitation as a way to decrease the number of addicts and, in turn, lower demand. In this and other cases, the relation between objectives and policy alternatives rests on causal assumptions that may prove to be questionable or plainly mistaken.

TABLE 5.4

Contrasts between Goals and Objectives

Characteristic	Goals	Objectives
Aims	1. Broadly stated (to upgrade the quality of health care. . . .)	1. Concrete (to increase the number of physicians by 10 percent. . . .)
Definitions	2. Formal (the quality of health care refers to accessibility of medical services. . . .)	2. Operational (the quality of health care refers to the number of physicians per 100,000 persons. . . .)
Time	3. Unspecified (in the future . . .)	3. Specified (in the period 2005–2015. . . .)
Measurement	4. Nonquantitative (adequate health insurance. . . .)	4. Frequently quantitative (the number of persons covered per 1,000 persons. . . .)
Target Groups	5. Broadly defined (persons in need of care. . . .)	5. Specifically defined (families with annual incomes below $19,000. . . .)

TABLE 5.5

Ten Tasks in Conducting a Cost–Benefit Analysis

Task	Description
Problem structuring	Formulation of metaproblem by defining boundaries of goals, objectives, alternatives, criteria, target groups, costs, and benefits
Specification of objectives	Conversion of general aims (goals) into temporally specific and measurable aims (objectives)
Specification of alternatives	Selection of most promising policy alternatives from a larger set of potential solutions defined at the problem structuring phase
Information search, analysis, and interpretation	Location, analysis, and interpretation of information needed to forecast the outcomes of specified policy alternatives
Identification of target groups and beneficiaries	Listing of all groups (stakeholders) that are a target of action (e.g., regulation) or inaction (status quo), or which will benefit from action or inaction
Estimation of costs and benefits	Estimation in units of monetary value of the specific benefits and costs of each alternative in all classes (internal and external, directly and indirectly measurable, primary and secondary, efficiency and redistributional)
Discounting of costs and benefits	Conversion of monetary costs and benefits into their present value on the basis of a specified discount factor
Estimation of risk and uncertainty	Use of sensitivity and *a fortiori* analysis to estimate probabilities that benefits and costs will occur in future
Choice of decision criterion	Choice of criterion for selecting among alternatives: Pareto improvement, net efficiency improvement, distributional improvement, internal rate of return
Prescription	Selection of alternative that is most plausible, considering rival ethical and causal hypotheses

Identification of alternative solutions. Once objectives have been specified, assumptions about the causes of a problem and its potential solutions are almost inevitably transformed into alternative policies to achieve these objectives. The way a problem has been structured—for example, the problem of cocaine may be formulated as a demand problem requiring the interdiction of the flow of drugs from South America—governs policies perceived to be appropriate and effective. If the problem search has been overly restrictive, inadvertently or for ideological reasons, the policy metaproblem will exclude otherwise relevant alternatives.

Information search, analysis, and interpretation. The task here is to locate, analyze, and interpret information relevant for forecasting the outcomes of policy alternatives. At this point, the principal objects of a forecast are the costs and benefits of policy alternatives already identified at the previous stage of analysis. Here, information may be acquired from available data on costs and benefits of similar existing programs. For example, prior to the establishment of the Drug Enforcement Administration, published budgetary data on the costs of drug enforcement activities of the U.S. Customs Service were used to estimate some of the costs of new drug interdiction policies. The benefits of interdiction in the form of decreased supply and demand were based on economic and administrative assumptions that eventually proved to be questionable.[27]

Identification of target groups and beneficiaries. Here the task is to conduct a stakeholder analysis which lists all groups that have a stake in the policy issue because they will be affected, negatively or positively, by the adoption and implementation of a policy. Target groups are the objects of new regulations or restrictions that typically involve the loss of freedom or resources—for example, new tax schedules that increase the tax burden on the middle class. By contrast, beneficiaries are groups that will gain from the adoption and implementation of prescriptions, for example, commercial truckers who receive increased net income as a consequence of some forty states moving from the 55 mph to the 65 mph speed limit in the late 1980s.

Calculation of costs and benefits. This task, perhaps the most difficult in CBA, requires the calculation in monetary terms of all benefits and costs that are likely to be experienced by target groups and beneficiaries. As we have seen (Figure 5.5), there are multiple types of benefits and costs: internal and external, directly and indirectly measurable, primary and secondary, net efficiency and redistributional. In many areas of public policy, it is difficult to estimate costs or benefits. For example, the monetary benefits of a fatality averted through mandatory seat belt laws, obligatory state vehicle inspections, or breast cancer screening programs are subject to broad disagreement, especially when the costs of these fatality-averting policies are multiplied by the value of a human life. The validity, reliability, and appropriateness of such measures are often disputed.

Discounting costs and benefits. If a certain level of costs and benefits is projected for a future time period, estimates must adjust for the decreasing real value of money due to inflation and future changes in interest rates. The real value of costs and benefits is usually based on the procedure of discounting, which projects future costs and benefits in terms of their present value. The *present value* of a future amount of money is the amount that, if the money were invested today, will grow to be as large as that future amount when the interest it will earn is taken into account. For example, the present value of $10,000 five years from now is the sum of $10,000 divided by 1.0 plus the rate of interest, say 4 percent, that could be earned on a sum of money over a 5-year period ($\$10{,}000 / 1.04^5 = \$8{,}219.27$). Sometimes, the value of a human life lost to traffic accidents is expressed as an hourly wage rate for a person multiplied by life expectancy of 76 years. If the human life being valued is that of a

[27]See George M. Guess and Paul G. Farnham, *Cases in Public Policy Analysis* (New York: Longman, 1989), Chapter 5, "Coping with Cocaine," pp. 7–48; and Constance Holden, "Street-Wise Crack Research," *Science* 246 (December 15, 1989): 1376–1381. The Bush and Obama administrations made an effort to *unsolve* this problem (Chapter 2).

20-year old who earns the minimum wage—and if the discount factor used is 4 percent—the present value of that person's life can amount to no more than a few thousand dollars. This raises ethical questions about the comparability of values attached to the lives of young and old, poor and rich, rural and urban populations.

Estimation of risk and uncertainty. The task here is to employ sensitivity analysis, a generic term that refers to procedures that test the sensitivity of conclusions to alternative assumptions about the probability of occurrence of different costs and benefits, or to different discount factors. It is difficult to develop reliable estimates because forecasts of the same future outcome often may have differing levels of accuracy that stem from differences in institutional, temporal, and historical contexts.

Choice of a decision criterion. Here the task is to specify a decision criterion or rule for choosing between two or more alternatives with different mixes of costs and benefits. These criteria, as we saw in the first part of this chapter, are of six types: efficiency, effectiveness, adequacy, equity, responsiveness, and appropriateness. Among efficiency criteria are net efficiency improvement (the net present value after discounting costs and benefits to their present value must exceed zero) and internal rate of return (the rate of return on a public investment must be greater than that which could be earned at the actual rate of interest paid). Effectiveness criteria include the marginal effectiveness of graduated levels of public investment in producing the greatest volume of some valued good or service (e.g., access to medical treatment per dollar invested in health maintenance organizations versus traditional health care providers). Distributional and redistributional criteria, respectively, include Pareto improvement (at least one person or group gains while none loses) and Rawlsian improvement (persons or groups worst off become better off). The choice of a decision criterion has important ethical implications, because decision criteria are grounded in different conceptions of moral obligation.

Prescription. The final task in cost–benefit analysis is to make a prescription by choosing among two or more policy alternatives. The choice of alternatives is seldom unequivocal, which calls for a critical analysis of the plausibility of prescriptions, taking into account rival causal and ethical hypotheses that may weaken or invalidate a prescription. For example, a prescription based on Pareto improvement may be weakened by establishing that the rich will benefit at the expense of the middle and lower classes. Viewed from the standpoint of income distribution, this has been the history of the United States and Europe over the past 80 years.[28]

Cost–Effectiveness Analysis

Cost–effectiveness analysis (CEA) is an approach to prescription that permits analysts to choose among two or more policies by quantifying their total costs and total effects. In contrast to CBA, which attempts to measure all costs and all benefits in a common unit of value, CEA uses two different units of value. Costs are measured in monetary units, while effectiveness is typically measured in units of goods, services, or some other valued effect.

[28]See Figure 5.2.

In the absence of a common unit of value, CEA does not permit net effectiveness or net benefit measures, because it is illogical to perform subtraction without a common unit of measure. It is possible, however, to produce cost–effectiveness ratios, for example, ratios of costs to units of health service delivered, for example, $1,200,050 total costs for 500 radiological exams. Here we calculate $1,200,050 / 500 = $2400.10 per exam.

These ratios are different from cost–benefit ratios. Whereas cost–effectiveness ratios tell us how many dollars are expended per unit produced, benefit–cost ratios tell us how many times more benefits than costs are produced in a given instance. Benefit–cost ratios must be greater than 1 if there are any net benefits at all. For example, if benefits are $4 million and total costs are $6 million, then the benefit–cost ratio is 4 to 6, or 0.67, and net benefits are minus $2 million. If net benefits are zero ($4 million minus $4 million equals zero), then the benefit–cost ratio is always 1 (4/4 = 1). By definition, then, a benefit–cost ratio must exceed 1 for there to be any net benefits. This is not true of effectiveness–cost ratios, which have a different meaning from one case to the next.

CEA, like its cost–benefit counterpart, may be applied prospectively (*ex ante*) as well as retrospectively (*ex post*). In contrast to cost–benefit analysis, cost–effectiveness analysis grew out of work done for the Defense Department in the early 1950s, and not from the work of Gallatin, Dupuis, and the U.S. Army Corps of Engineers. Much of the post-World War II development of CEA was carried out by the RAND Corporation in projects designed to estimate the costs of alternative military strategies and weapons systems. In the same period, CEA was applied to problems of program budgeting in the Department of Defense and, beginning with the 1960s, was extended to other federal agencies.[29]

CEA is particularly appropriate for questions involving the most efficient way to use resources to attain objectives that cannot be expressed in monetary terms. Cost–effectiveness analysis has been used to prescribe alternative policies and programs in criminal justice, manpower training, transportation, health, defense, and other areas.

Cost–effectiveness analysis, when used to make prescriptions in the public sector, has several distinguishing characteristics:

1. CEA, because it avoids problems of measuring benefits in monetary terms, is sometimes more easily applied than cost–benefit analysis.
2. CEA epitomizes *technical rationality,* because it attempts to determine the utility of policy alternatives but without relating their consequences to global economic efficiency or aggregate social welfare.
3. CEA, because it relies minimally on market prices, is less dependent on the logic of profit maximization in the private sector. Using cost–effectiveness analysis, for example, it is sometimes difficult to determine whether benefits exceed costs or whether alternative investments in the private sector would have been more profitable.

[29]For documentation on the early years, see David Novick, *Efficiency and Economy in Government through New Budgeting and Accounting Procedures* (Santa Monica, CA: RAND Corporation, February 1954); Roland J. McKean, *Efficiency in Government through Systems Analysis* (New York: Wiley, 1958); Charles J. Hitch and Roland J. McKean, *The Economics of Defense in the Nuclear Age* (New York: Athenaeum, 1965); and T. A. Goldman, ed., *Cost-Effectiveness Analysis* (New York: Frederick A. Praeger, 1967).

4. CEA is well suited to the analysis of externalities and intangibles, because these types of effects are often difficult to express in dollars as a common unit of measure.
5. CEA typically addresses fixed-cost (Type I) or fixed-effectiveness (Type II) problems, whereas cost–benefit analysis typically addresses variable-cost-variable-effectiveness (Type III) problems.

The strengths of CEA are its comparative ease of application, its capacity to treat collective and quasi-collective goods whose values cannot be estimated on the basis of market prices, and its appropriateness for analyzing externalities and intangibles.

The main limitation of CEA is that prescriptions are less easily related to questions of aggregate social welfare. Unlike CBA, efforts to measure costs and effectiveness are confined to given programs, jurisdictions, or target groups and cannot be used to calculate net income benefits as a measure of the aggregate satisfaction experienced by members of a community.

CEA also attempts to consider all costs and benefits of policies and programs, with the exception that benefits are not measured in monetary terms. Many of the same types of costs and benefits as those discussed previously are important in CEA: inside versus outside costs and benefits, directly measurable versus indirectly measurable costs and benefits, and primary versus secondary costs and benefits. While it is not possible to calculate net efficiency improvement, redistributional effects may be analyzed.

The tasks in conducting a CEA are similar to those required in CBA (Table 5.5), with two exceptions. Only costs are discounted to their present value and criteria of adequacy differ from those normally used in CBA. In CEA, four criteria of adequacy are often employed:

1. *Least-cost criterion.* After establishing a fixed level of effectiveness, the costs of programs with approximately equal effectiveness are compared. Programs which meet the fixed level of effectiveness at least cost are prescribed.
2. *Maximum-effectiveness criterion.* After establishing an upper level of permissible costs (usually a budgetary constraint), programs with equal costs are compared. The program that meets the fixed-cost level with the maximum effectiveness is prescribed.
3. *Marginal effectiveness.* If units of some service or good and its costs can be expressed on two continuous scales, the marginal effectiveness of two or more alternatives can be calculated. For example, the costs of providing security services by municipal police and by private security firms can be expressed on two continuous scales: a graduated scale of costs associated with each hour of patrols and the number of crimes against property reported, deterred, investigated, and cleared. A continuous cost–effectiveness function may be established for each type of service provider. The provider with the highest effectiveness–cost ratio at any point along the function beyond some minimum level of effectiveness has the greater marginal effectiveness—that is, the highest effectiveness achieved at the last dollar expended (see point $E_1 - C_1$ in Figure 5.1).
4. *Cost–effectiveness.* The cost–effectiveness of two or more alternatives is the cost per unit of goods or services. For example, the 55 mph and 65 mph speed limits in the United States involve different costs per fatality averted, with the latter ranking

slightly lower in cost–effectiveness. If the 55 mph and 65 mph speed limits have approximately equal costs (the denominator), but the latter averts slightly fewer fatalities (the numerator), then the average cost per fatality averted will be greater for the 65 mph speed limit.

METHODS OF PRESCRIPTION

A range of methods are available for making prescriptions when using CBA and CEA (Table 5.6). These methods are most easily understood if they are viewed as tools for carrying out tasks of CBA described in Table 5.5. Cost–benefit and cost–effectiveness analysis, as we saw earlier, differ in certain important respects. These differences, however, are mainly related

TABLE 5.6

Methods of Prescription

Task	Method/Technique
Problem structuring[(1)]	Boundary analysis
	Classificational analysis
	Hierarchy analysis
	Multiple perspective analysis
	Argumentation analysis
	Argumentation mapping
Specification of objectives	Objectives mapping
	Value clarification
	Value critique
Information search, analysis, and interpretation	Boundary analysis[(2)]
Identification of target groups and beneficiaries	Boundary analysis
Estimation of costs and benefits	Cost element structuring
	Cost estimation
	Shadow pricing
Discounting of costs and benefits	Discounting
Estimation of risk and uncertainty	Feasibility assessment[(3)]
	Constraints mapping
	Sensitivity analysis
	A fortiori analysis
Choice of decision criterion	Value clarification
	Value critique
Prescription	Plausibility analysis

Notes:
[(1)] See Chapter 3 for these methods.
[(2)] See Chapter 3.
[(3)] See Chapter 4.

to the choice of decision criteria—for example, net efficiency improvement cannot be estimated with CEA because benefits are not expressed in monetary terms. With few exceptions (e.g., calculating the present value of future benefits), methods of CBA are also appropriate for CEA.

Objectives Mapping

A frequent difficulty experienced in making policy prescriptions is knowing what objectives to achieve. *Objectives mapping* is a technique used to array goals and objectives and their relationship to policy alternatives. Goals, objectives, and alternatives that have been identified by one or more methodologies of problem structuring (Chapter 3) may be depicted in the form of an *objectives tree,* which is a pictorial display of the overall structure of objectives and their relationships.[30] The objectives, when mapped in the form of a tree diagram, often form a hierarchy in which certain objectives that are necessary for the attainment of other objectives are arranged vertically. An objectives tree for the development of a national energy policy is illustrated in Figure 5.6.

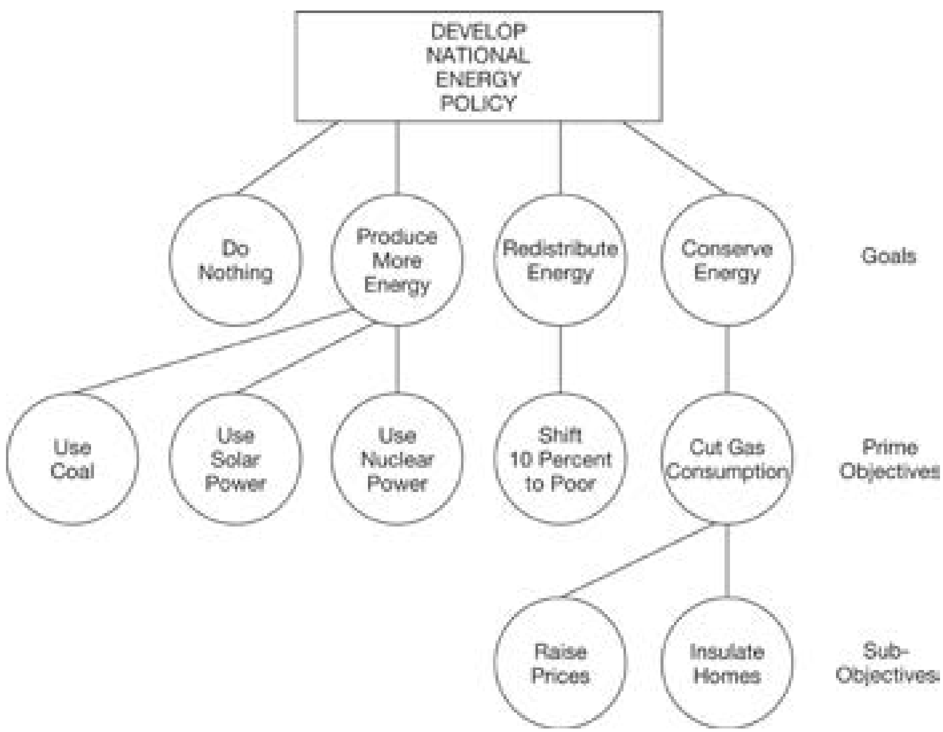


FIGURE 5.6
Objectives Tree for National Energy Policy

[30]See T. Harrell Allen, *New Methodologies in Social Science Research: Policy Sciences and Futures Research* (New York: Frederick A. Praeger, 1978), pp. 95–106.

In constructing an objectives tree, existing or new objectives may be generated. If it is necessary to generate new objectives, there are several problem-structuring techniques available for this purpose, including hierarchy analysis, classificational analysis, brainstorming, and argumentation analysis (see Chapter 3). Once a set of objectives has been obtained, it is possible to arrange them hierarchically in a tree such as that of Figure 5.6. Objectives trees are more general at the top and progressively more detailed as one moves lower in the hierarchy. Typically, the uppermost part of an objectives tree contains broad purposes, while lower levels represent goals, prime objectives, and subobjectives. Observe that when we read downward, we answer the question "*How* should we achieve this objective?" When we read upward, we answer the question "*Why* should we achieve this objective?" Objectives mapping is useful, not only for purposes of mapping the complexities of policy implementation (i.e., "how" questions), but also for clarifying the ends of action (i.e., "why" questions). Most objectives can be regarded as both ends and means.

Values Clarification

Values clarification is a procedure for identifying and classifying values that underlie the selection of policy objectives. The need for values clarification in making policy prescriptions is most evident when we consider competing criteria for prescription (effectiveness, efficiency, adequacy, responsiveness, equity, appropriateness) and the multiple forms of rationality with which these criteria are associated. Values clarification helps answer such questions as the following: What values underlie the choice of policy objectives? Are these values those of policy analysts, of policymakers, of particular social groups, or of society as a whole? What circumstances explain why certain groups are committed to these values and objectives, while others are opposed? Finally, what reasons are offered to justify particular values, premises and objectives?

There are several steps in values clarification:

1. Identify objectives of a policy or program. Objectives may simply be listed or displayed in the form of an objectives tree (Figure 5.6).
2. Identify stakeholders who affect and are affected by the attainment or non-attainment of the objectives. Be sure to include policy analysts and advisers in the list of stakeholders.
3. List the value premises that underlie each stakeholder's commitment to objectives. For example, environmentalists may value energy conservation because it preserves the esthetic qualities of the natural environment, while suburban homeowners may value energy conservation because it provides for a more secure material existence.
4. Classify values into those that are simply expressions of personal taste or desire (*value expressions*), those that are statements about the beliefs of particular groups (*value statements*), and those that are judgments about the universal goodness or badness of the actions or conditions implied by the value (*value judgments*). For example, you as policy analyst may express a personal desire for more energy consumption, environmental groups may state the values that motivate their belief in energy conservation, and oil producers may offer value judgments about increased energy production that are based on beliefs in universal rights to use private property.

5. Further classify value premises into those that provide a *basis for explaining* objectives (e.g., environmentalists seek to conserve energy because this is consistent with their belief in the inviolability of nature) and those that provide a *ground for justifying* objectives (e.g., energy conservation is an appropriate objective because nature and humanity alike are entitled to rights of self-protection).

The advantage of values clarification is that it enables us to go beyond the analysis of objectives as if they were merely expressions of personal desire or taste. Values clarification also takes us one step beyond the explanation of circumstances that explain objectives. While the *bases* of values are important—for example, it is important to know that the poor favor greater social equity—ethical and moral disputes cannot be resolved without examining the *grounds* for justifying objectives.

Values Critique

Values critique is a set of procedures for examining the persuasiveness of conflicting arguments offered in the course of a debate about policy objectives. Whereas values clarification enables us to classify values according to their form of expression, the standard contexts in which values are held, and their explanatory and justificatory functions, values critique allows us to examine the role of values in policy discourse. Values clarification focuses on the objectives and underlying values of individual stakeholders. By contrast, values critique focuses on conflicts among the objectives and underlying values of different stakeholders. Values clarification has a static descriptive quality, while values critique looks toward changes in values that may result from reasoned policy discourse.

Procedures for values critique are an extension of the ethical mode of policy argument discussed in Chapter 8. In conducting a value critique, it is necessary to

1. Identify one or more claims that set forth a prescription for action.
2. List all stakeholders who will affect and be affected by the implementation of the prescription.
3. Describe each stakeholder's argument for and against the prescription.
4. Identify each element in the debate: information (*I*), claim (*C*), qualifier (*Q*), warrant (*W*), backing (*B*), and rebuttal (*R*).
5. Assess the *ethical persuasiveness* of each argument and determine whether to retain, alter, or reject the prescription.

To illustrate the process of value critique, imagine a policy analyst who has completed a cost–benefit analysis of home insulation. The prescription or claim (*C*) is that the government should adopt the home insulation program. Information (*I*) indicates that the net benefits of the program are $50 million and the benefit–cost ratio is 5:1. As the program will result in considerable energy savings, it is almost certainly justified (*Q*). But after the arguments of each stakeholder have been analyzed, the prescription seems less persuasive (Figure 5.7). The original warrant (*W*) used to support the prescription is that of increased social welfare, a warrant that in turn might be backed (*B*) by the Pareto or Kaldor–Hicks

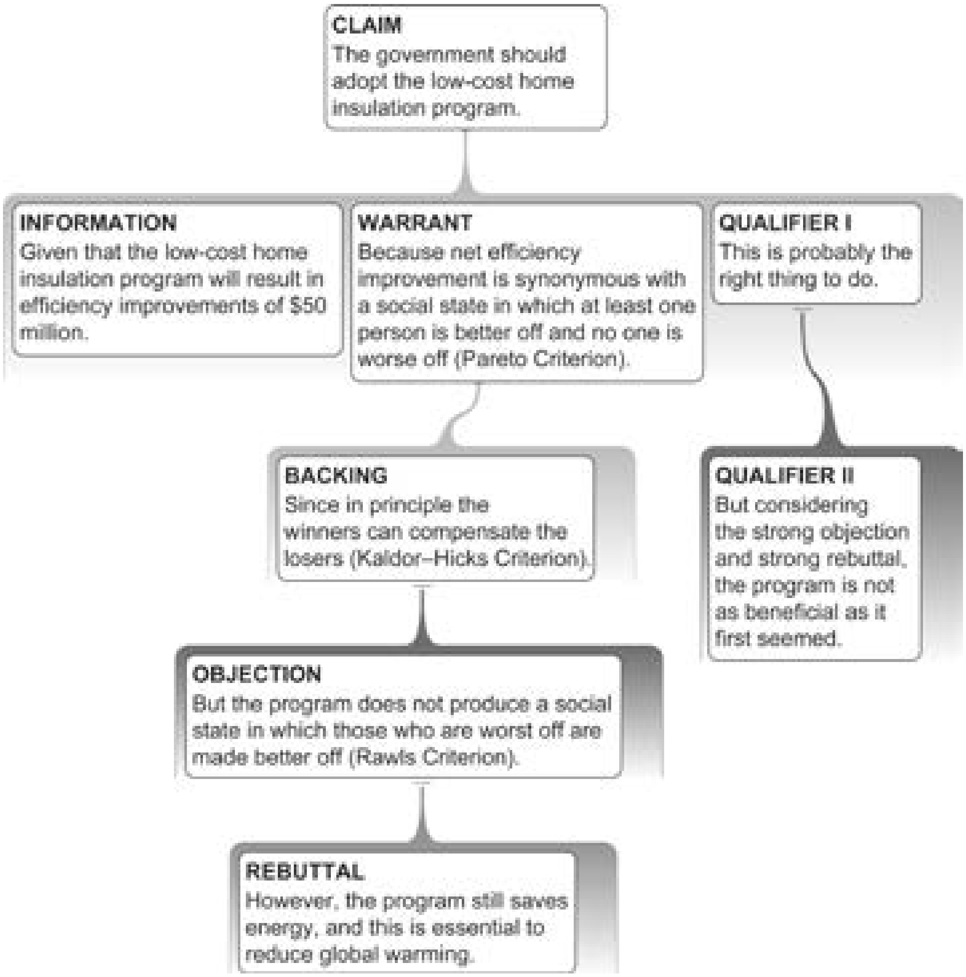


FIGURE 5.7
Value-Critical Discourse

criterion. The rebuttal (*R*) is that the program will not benefit the poor and the elderly. This rebuttal is backed by the Rawls criterion that persons who are worst off should gain.

This instance of value-critical discourse is simultaneously economic and ethical, because it involves questions of efficiency and justice at the same time. One result of the debate might be to modify the prescription so that a government would provide low-cost home insulation on a graduated-cost basis. Those who cannot afford it might pay nothing, while those in higher-income brackets would pay more, in graduated steps. Yet this would make the program less attractive to those who, in consuming the most heating fuel, would pay more of the costs. If equal costs were maintained among these groups across the board (with the exception of the poor and the elderly, who would pay nothing), the total net benefits may well be negative.

While value critique cannot finally answer the question of how much efficiency should be sacrificed for a given increase in social equity, it does make possible reasoned ethical discourse about such questions, rather than expect cost–benefit analysis to answer questions for which it is ill suited. Economics is not ethics any more than price is value.[31]

Cost Element Structuring

Opportunity costs, as we have seen, are the benefits that could be obtained from investing in a policy if we did not invest in another policy. In other words, costs are benefits forgone or lost, including net benefits (benefits minus costs).

Cost element structuring is a procedure for classifying and describing costs that will be incurred by adopting and operating a program.[32] The product of cost element structuring is a list of activities, equipment, and supplies that require the expenditure of funds. The list

TABLE 5.7

Cost Element Structure

I. Primary (direct) costs
  1. One-time fixed costs
     - Research
     - Planning
     - Development, testing, and evaluation
  2. Investment costs
     - Land
     - Building and facilities
     - Equipment and vehicles
     - Initial training
  3. Recurring (operating and maintenance) costs
     - Salaries, wages, and fringe benefits
     - Maintenance of grounds, vehicles, and equipment
     - Recurrent training
     - Direct payments to clients
     - Payments for extended support services
     - Miscellaneous materials, supplies, and services

II. Secondary (indirect) costs
  1. Costs to other agencies and third parties
  2. Environmental degradation
  3. Disruption of social institutions
  4. Other

[31]This is true because prices are only one kind of value.

[32]See H. G. Massey, David Novick, and R. E. Peterson, *Cost Measurement: Tools and Methodology for Cost-Effectiveness Analysis* (Santa Monica, CA: RAND Corporation, February 1972).

should be exhaustive, with categories that are mutually exclusive. If the list is exhaustive, no important costs will be overlooked. If the list contains mutually exclusive cost categories, *double-counting* will be prevented.

A cost element structure contains two main divisions: *primary* (direct) and *secondary* (indirect) costs. Primary costs are subdivided into several categories: *one-time fixed costs, investment costs,* and *recurrent* (*operating and maintenance*) *costs.*[33] A typical cost element structure is illustrated in Table 5.7.

Cost Estimation

Cost estimation is a procedure that provides information about the dollar value of items in a cost element structure. While the cost element structure indicates what elements should be measured, cost estimation actually measures them. A *cost-estimating relationship* is an explicit measure of the relationship between the quantity of activities, materials, or personnel and their costs. The simplest type of cost-estimating relationship is cost per unit, for example, the cost per acre of land, per square foot of facilities, or per worker. More complex cost-estimating relationships involve the use of regression analysis and interval estimation (Chapter 4). A linear regression may be used to provide an interval estimate of vehicle maintenance costs, using miles driven as the independent variable.

Cost-estimating relationships are central to the process of making prescriptions, because measures of the relations between the quantity of activities, materials, and personnel and their costs are essential for comparisons of alternatives. Cost-estimating relationships are used to construct cost models that represent some or all of the costs that will be required to initiate and maintain a program. A simplified cost model is illustrated in Figure 5.8.

Shadow Pricing

Shadow pricing is a procedure for making judgments about the monetary value of benefits and costs when market prices are unavailable or distorted. Market prices are distorted for a number of reasons, including unfair competition, monopolistic or oligopolistic practices, and government price supports. In such cases, upward (or downward) adjustments can be made in the actual market price of a good or service. In other cases, market prices are simply unavailable. This is frequently the case with certain collective goods (e.g., clean air or a sense of community) that we discussed earlier as *intangibles.* There are several ways to estimate the price of intangibles.[34]

1. *Comparable prices.* Prices for comparable or similar items in the market may be used to estimate the price of intangibles. For example, in estimating the economic benefits of time saved by commuting on a new mass rapid transit system, we might use area wage rates as a measure of the monetary value of time saved by using the new transportation facilities.

[33]Poister, *Public Program Analysis,* pp. 386–387, and H. P. Hatry, R. E. Winnie, and D. M. Fisk, *Practical Program Evaluation for State and Local Government* (Washington, DC: Urban Institute, 1973).
[34]Poister, *Public Program Analysis*, pp. 417–419.

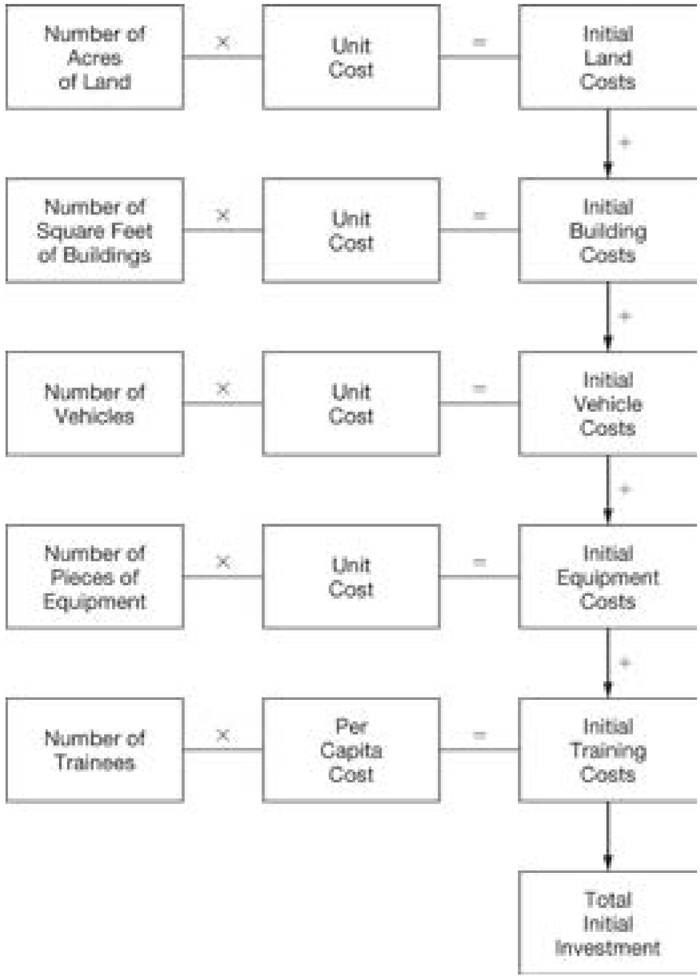


FIGURE 5.8
Simplified Partial Cost Model for Total Initial Investment

2. *Consumer choice.* Here the analyst estimates the value of intangibles such as travel time by observing consumer behavior in situations where individuals are forced to choose between a given intangible and money. For example, estimates of the value of time can be made by observing choices made between low-cost and time-consuming versus high-cost and time-reducing modes of travel. This may help to determine how much consumers are willing to pay for reduced travel time, assuming that consumers have complete information about time and costs and that both alternatives are equal in all respects except time and cost.

3. *Derived demand.* Here the value of intangibles for which there are no market prices (e.g., satisfaction with government parks and recreation areas that charge no fees) can be estimated on the basis of the indirect costs paid by visitors. A demand curve (i.e., a curve that displays the various price–quantity combinations at which consumers will use a service) can be derived from the costs paid by consumers in traveling to the park. These indirect costs are taken to be actual prices paid for the use of parks and recreation areas. This procedure assumes that the sole purpose of the travel is to use the park or recreation area.
4. *Survey analysis.* Here we survey citizens by interview or mailed or electronic questionnaire. Respondents indicate at what level of costs they will be willing to pay for a given service (e.g., bus transportation). One weakness of this procedure is the so-called *free-rider problem*, that is, a situation where consumers will falsely claim that they are willing to pay for a service in the hope that they will benefit from the service while others pay for it.
5. *Cost of compensation.* Here the analyst estimates the value of intangibles—particularly those that occur in the form of negative externalities such as pollution—by obtaining prices for actions required to correct them. For example, the costs of environmental damage can be estimated by basing dollar values on the prices of programs for water purification, noise abatement, or reforestation. Similarly, the benefits of an antipollution enforcement program might be estimated on the basis of medical costs that will be avoided because people contract less lung cancer, emphysema, and other chronic diseases.

Constraints Mapping

Constraints mapping is a procedure for identifying and classifying obstacles that stand in the way of achieving policy objectives. Generally, constraints fall into six categories:

1. *Physical constraints.* The attainment of objectives may be limited by the state of development of technology. For example, the reduction of pollution through the use of solar energy is constrained by the level of development of solar technology.
2. *Legal constraints.* Public law, property rights, and agency regulations may limit attempts to achieve objectives. For example, social programs designed to redistribute resources to the poor are constrained by reporting requirements.
3. *Organizational constraints.* The organizational structure and processes available to implement policies may limit efforts to achieve objectives. For example, excessive centralization, poor management, and low morale constrain the effectiveness of programs and policies.[35]

[35]For a general treatment of these organizational constraints, see Vincent Ostrom, *The Intellectual Crisis in American Public Administration* (Tuscaloosa, AL: University of Alabama Press, 1974).

4. *Political constraints.* Political opposition imposes limitations on the initial acceptance and the implementation of policies. Such opposition is reflected in organizational inertia and tendencies to avoid problems by practicing incremental decision-making. For example, consumer protection, environmental protection, and energy conservation have taken years before they were placed on the agenda of legislatures.[36]
5. *Distributional constraints.* Public programs designed to provide social services often must ensure that benefits and costs are equitably distributed among different groups. Programs that achieve higher net efficiency benefits frequently produce lower social equity, and vice versa.
6. *Budgetary constraints.* Government budgets are limited, requiring that expenditures be considered in light of scarce resources. Fixed budgets create Type I problems, which force the consideration of alternatives that maximize effectiveness only within the limits of available resources.

An effective way to identify and classify constraints is to construct a *constraints tree,* which is a graphic display of obstacles that stand in the way of achieving objectives. A constraints tree involves the superimposition of constraints on an objectives tree (Figure 5.6). A simplified constraints tree for national energy policy is illustrated in Figure 5.9.

Cost Internalization

Cost internalization incorporates outside costs (externalities) into the internal cost element structure. Costs as well as benefits can be internalized by incorporating positive and negative spillovers of public policies and programs. When costs are fully internalized, there are no externalities because, by definition, all costs are internal. This means that a variety of important external costs—for example, costs of pollution, environmental degradation, or social dislocation—are explicitly built into the cost element structure.

In general, there are four kinds of spillovers or externalities[37]:

1. *Production-to-production spillovers.* The products of a program serving one target group or jurisdiction may affect (positively or negatively) the products of another. For example, successful alcoholism and drug-treatment programs may result in a decrease of law enforcement activities in the surrounding community.
2. *Production-to-consumption spillovers.* The products of a particular program may affect the quality and quantity of goods consumed by members within another target group or jurisdiction. For example, publicly funded highway projects may displace residents from their homes or change the demand for goods and services in neighboring areas.

[36]See, for example, Roger W. Cobb and Charles D. Elder, *Participation in American Politics: The Dynamics of Agenda-Building* (Baltimore, MD: Johns Hopkins University Press, 1983), and Paul A. Sabatier and Christopher M. Weible, "The Advocacy Coalition Framework: Innovations and Clarifications," in P. A. Sabatier (ed.), *Theories of the Policy Process, 2nd ed.* (Boulder, CO: Westview Press, 1999), pp. 189–217.

[37]Hinrichs and Taylor, *Systematic Analysis*, pp. 18–19.

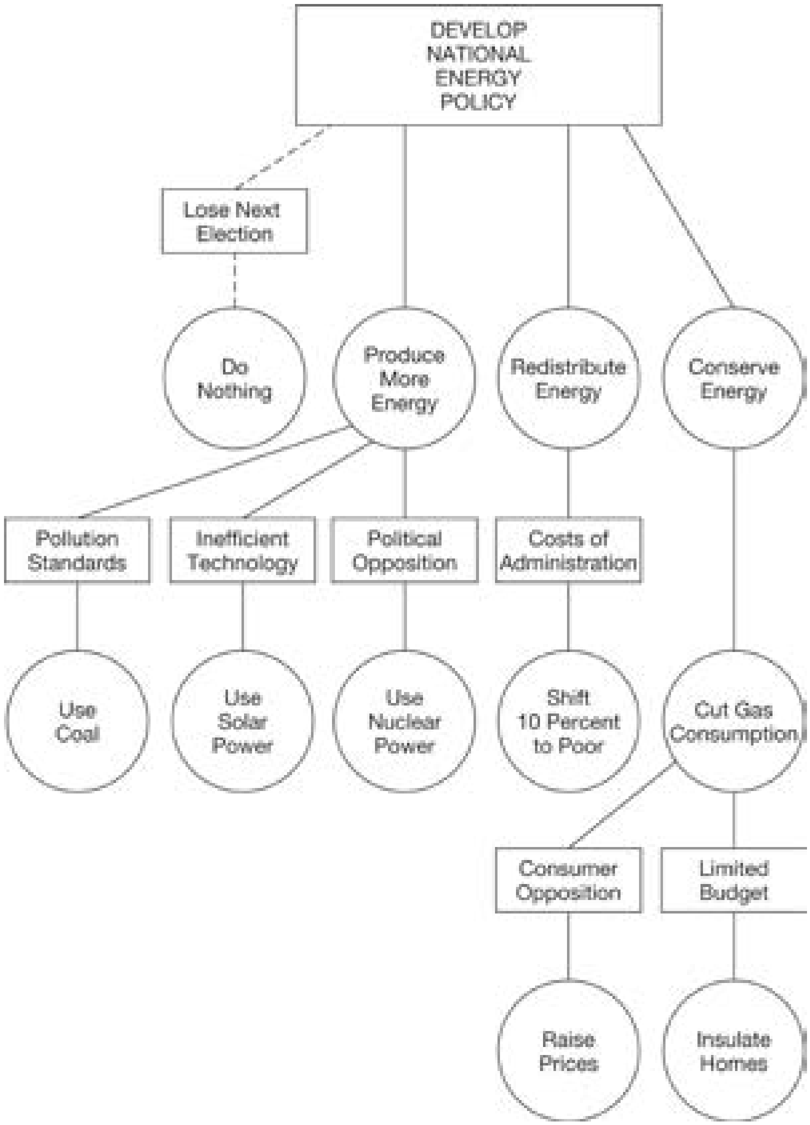


FIGURE 5.9
Constraints Map for National Energy Policy

3. *Consumption-to-consumption spillovers.* The consumption activities of public programs in one area may affect consumption within adjacent target groups or jurisdictions. For example, the construction of a large government office facility may make it impossible for local citizens to find adequate parking.

4. *Consumption-to-production spillovers.* The consumption activities in one area may affect production activities in adjacent areas. For example, the construction of a public housing project may improve the market for local businesses.

The importance of cost internalization is evident in Table 5.8, which shows total, outside, and inside costs for three programs designed to provide maternal care. If all three programs are equally effective, an analyst at the state level who has not internalized federal cost sharing will prescribe Program I, because it satisfies the least-cost criterion. If federal costs are internalized, Program III will be prescribed, because total costs are least. This example calls attention to the relative nature of internal and external costs, because what is an externality to one party is not to another.

Discounting

Discounting, as we saw in the listing of steps in a cost–benefit analysis (Table 5.5), is a way to estimate the present value of costs and benefits that will be realized in the future. Discounting is a way to take into account the effects of time. Time is important because future costs and benefits have less value than present ones. A dollar spent today is more valuable than a dollar spent 1 or 2 years from now, because today's dollar can be invested so that it will generate $1.04 one year from now (at 4% interest). Apart from interest earnings, many people prefer to consume now rather than later, since the future is uncertain and consumption now will produce immediate satisfaction.

Policies and programs generate different levels of costs and benefits over time. These flows of costs and benefits, called *cost (and benefit) streams,* are unevenly distributed in time. For this reason, each cost and benefit stream must be discounted to its present value. For example, two health programs with equal effectiveness and identical total costs over a 5-year period would appear to be equally effective. Yet in one of these programs (Program I), costs are heavily concentrated in the first year, while in the other (Program II), they are concentrated in the last year. If we assume that the dollar is losing value at the rate of 10 percent per year, then Program II achieves the same level of effectiveness at least cost.

TABLE 5.8

Internal, External, and Total Costs of Maternal Care

	Program					
Type of Cost	Maternal and Infant Care Project I		Neighborhood Health Center II		Private Physician III	
Cost to state (internal)	$31	10%	$96	48.3%	$85	48.3%
Cost to federal government (external)	282	90	103	51.7	90	51.7
Total cost	$313	100%	$199	100%	$175	100%

Source: Adapted from Burt (1974): 35.

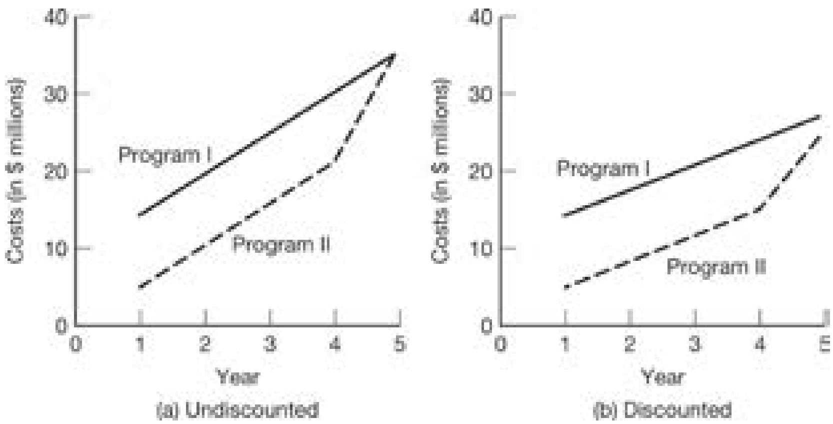


FIGURE 5.10
Comparison of Discounted and Undiscounted Costs Cumulated for Two Programs with Equal Effectiveness

If we do not take time and the changing value of money into account, there is no basis for distinguishing the two programs (Figure 5.10).

The present value of a future benefit or cost is found by using a *discount factor,* which expresses the amount by which the value of future benefits and costs should be decreased to reflect the fact that present dollars have more value than future ones. Once we know the discount factor, we simply multiply this factor by the known dollar value of a future benefit or cost. Given a certain *discount rate* (sometimes called the *social discount rate*), that is, the rate at which the value of future benefits and costs should be reduced, the *discount factor* may be calculated by knowing the number of years over which future benefits and costs must be discounted. The formula for the discount factor is the reciprocal of compound interest:

$$\mathrm{DF} = \frac{1}{(1+r)^n}$$

where r is the discount rate and n is the number of years over which benefits or costs are discounted.[38] For example, to calculate the discount factor (DF) for a discount rate of 10 percent at 5 years, we make the following computations:

[38]The discount factor assumes that future dollar values in a cost (or benefit) stream are unequal (for example, heavy costs at the beginning of a period decrease over time). If the future dollar values in a cost (or benefit) stream are identical from year to year (e.g., $1 million in costs every year for 5 years), it is customary to use a shortcut, called an *annuity factor* (annuity means annually, or regularly over a number of years). The formula for the annuity factor is

$$\begin{aligned} DF &= \frac{1}{(1+r)^n} \\ &= \frac{1}{(1+0.10)^5} = \frac{1}{(1.1)^5} \\ &= \frac{1}{(1+0.10)^5} = \frac{1}{(1.1)^5} \\ &= \frac{1}{(1.61)} \end{aligned}$$

If we want to discount the value of $100 in costs incurred in the 5th year, we simply multiply this cost by the discount factor, that is, $100 × 0.621 = $62.10. In other words, the present value of $100 in costs incurred in the 5th year at a discount rate of 10 percent is $62.10. *Present value* is the current dollar value of future costs or benefits that have been multiplied by the appropriate discount factor. The formula for the present value of a stream of costs or benefits is

$$PV = \Sigma\left(FV \cdot DF\right)$$

where FV designates the future value of costs or benefits, DF is the discount factor, and the symbol Σ (sigma) tells us to sum or add up all the products of costs or benefits times the discount factor for all years in the cost (or benefit) stream.[39]

The calculation of the present value of a cost stream is illustrated in Table 5.9. Observe that the future value of all undiscounted costs is $180 million, while the present value of the discounted cost stream is $152.75 million. Assume that the present value of the *benefit* stream is $175 million. If we were using the net benefit criterion to make a choice (i.e., net benefits must be greater than zero), the undiscounted costs would result in a rejection of the program ($175–180 = –$5). By contrast, when we discount the cost stream, our prescription would be positive, because net benefits are $22.5 million ($175–152.75 = $22.5).

Once we have identified the appropriate discount rate, it is a simple matter to calculate the discount factor and the present value of a cost (or benefit) stream. The problem is that the choice of a discount rate depends on judgments that reflect the values of analysts, policymakers, and other stakeholders. For example, the higher the rate at which we discount future benefits, the more difficult it is to obtain net benefits and benefit–cost ratios that justify public programs. If discount rates are uniformly high, this leads to a minimal role for government investments and a maximal one for private ones. If, on the other hand,

---

$$AF = 1 - \left(\frac{1}{1+r}\right) n/r.$$

Stated more concisely, $AF = 1 - DF/r$.

[39]Another formula for present value, which gives identical results, is $PV = FV/(1 + r)^n$, where FV is future costs or benefits, r is the discount rate, and n is the number of years over which costs or benefits are discounted.

TABLE 5.9

Calculation of Present Value of Cost Stream at 10 Percent Discount Rate over 5 Years (in Millions of Dollars)

Year	Future Value (FV)	Discount Rate (r)	Number of Years (n)	Discount Factor (DF) [$1/(1 + r)^n$]	Present Value (PV) [FV.DF]
2010	$100	0.10	1	0.909	$90.90
2011	50	0.10	2	0.826	41.30
2012	10	0.10	3	0.751	7.51
2013	10	0.10	4	0.683	6.83
2014	10	0.10	5	0.621	6.21
	$180.0				$152.75

discount rates are uniformly low, this encourages an expanded role for government. For this reason, it is important to recognize competing bases for selecting a discount rate:

1. *Private discount rate.* Here the selection of a discount rate is based on the average rate of interest charged for borrowing money in the private sector. The argument for using private rates is that public investments are made with the tax monies of citizens who could have invested in the private sector. The private rate of discount therefore measures the benefits that would have been obtained if funds had been left in the private sector. These are the opportunity costs of public (versus private) investment. The arguments against private rates are that private rates differ widely (from 3 to 20 percent) because of market distortions; private rates do not reflect the external social costs of private investment (for example, pollution); and private rates reflect narrow individual and group preferences and not those of society as a whole.
2. *Social discount rate.* Here the selection of a discount rate is based on judgments about the social time preference of society as a whole. The *social time preference* refers to the collective value that society places on benefits or costs realized in some future time period. This collective value is not simply the sum of individual preferences, because it reflects a collective sense of what is valuable for the community as a whole. The *social rate of discount,* that is, the rate at which future collective benefits and costs should be discounted, is generally lower than the private rate. The arguments for a social rate of discount are that the social rate compensates for the narrowness and short-sightedness of individual preferences; it takes into account external social costs of private investment (for example, the depletion of finite natural resources); and it reflects a concern with the security, health, and welfare of future generations. Arguments against the social rate of discount are that it is economically inefficient or "irrational" because it should be *higher* than the private rate. Taking a number of considerations into account, Moore and colleagues propose a social discount rate (SDR) between 3.0 and 4.0 percent.[40] In the period following the recession of 2008, the appropriate rate lies closer to 3.0. The lower

[40]Mark A. Moore, Anthony E. Boardman, Aidan R. Vining, David L. Weimer, and David H. Greenberg, "'Just Give Me a Number!' Practical Values for the Social Discount Rate," *Journal of Policy Analysis and Management* 23, 4 (Autumn 2004), pp. 789–812.

the SDR, the higher the present value of future costs and benefits, and the greater the propensity to invest in the public sector.

3. *Government discount rates.* Here a discount rate is selected on the basis of the current costs of government borrowing. The problem is that the rate at which federal, state, and local governments borrow money varies considerably. While the Office of Management and Budget (OMB) has advocated varying discount rates for government programs, a standard rate of, say, 6 percent does not reflect the opportunity cost of investments when lower interest rates are available within a given state, region, or community.

The choice of a discount rate is closely connected to competing views of the proper role of government in society and to the different forms of rationality discussed earlier. Under these circumstances, analysts must assess the sensitivity of net benefits or benefit-cost ratios to alternative discount rates.

Another alternative, which avoids issues of the selection of a discount rate, is the *internal rate of return* on investment, which permits a choice among policies when the rate of undiscounted benefits returned for each discounted dollar of costs exceeds the actual rate at which investment funds may be borrowed. Here, as elsewhere, it is essential to examine the underlying assumptions and ethical implications of choice.

Sensitivity Analysis

Sensitivity analysis is a procedure for examining measures of costs and benefits under different assumptions. In comparing two or more alternatives, there is usually uncertainty about the expected outcomes of a policy, although a single overall measure of costs and benefits (e.g., net benefits) may have been calculated. In such circumstances, which are often characterized as a deterministic (as distinguished from probabilistic) choice, we can introduce alternative assumptions about future costs—for example, assumptions about the likelihood of high, medium, and low costs—and compute net benefits under each of these assumptions. The point is to see how "sensitive" the results are to different assumptions.

For example, in comparing two personnel training programs, we may examine the sensitivity of transportation costs to potential changes in the price of gasoline. Here we may introduce assumptions that the price of gasoline will increase by 10 percent, 20 percent, and 30 percent over the life of the program. As Program I is situated in a semirural area, trainees must drive a longer distance to reach the training site than trainees in Program II, which is located in the city. In the absence of a sensitivity analysis, a marked increase in gasoline prices could result in incorrectly prescribing the costlier of the two programs. An example of this kind of sensitivity analysis is provided in Table 5.10. Another type of sensitivity analysis, and one we already discussed in Chapter 4, involves the use of regression analysis to make interval estimates of the values of costs or benefits—in effect, "best" case and "worst" case scenarios.

Plausibility Analysis

Plausibility analysis enables the testing of a prescribed policy against rival policies. Plausibility analysis—where plausibility is understood as an informed subjective probability based on prior experience—may be used in analyzing rival claims. There are several types

TABLE 5.10

Sensitivity Analysis of Gasoline Prices on Costs of Training Programs

	Price Increase		
Program	*10%*	*20%*	*30%*
I. Semirural			
Total costs	$4400	$4800	$5200
Trainees	$5000	$5000	$5000
Costs per trainee	$0.88	$0.96	$1.04
II. Urban			
Total costs	$4500	$4800	$5100
Trainees	$5000	$5000	$5000
Costs per trainee	$0.90	$0.96	$1.02

of rival claims, each of which is a threat to the plausibility of choices based on CBA. These threats—which are analogous to rival hypotheses or "threats to the validity" of claims about the effects of policies on specific outcomes—are illustrated in Figure 5.11.

- *Invalidity.* A policy prescription is based on an invalid assumption about the causal relation between a policy and its outcomes. For example, debates about the efficacy of the 55 mph speed limit (National Maximum Speed Law of 1974) have focused on the extent to which enforcing a maximum speed on all intercity highways is responsible for the observed decline in traffic fatalities. The claim that the new maximum speed (55 mph) was responsible for the decline of 9,800 fatalities between 1973 and 1975 has been challenged on grounds that the decline was due to the new compressed range or standard deviation of speeds; improvements in automobile and highway safety; the interaction of maximum speed with population density; and the effects of recession, unemployment, and declining gasoline prices on miles driven and, consequently, fatality rates.[41]
- *Inefficiency.* Estimates of the net efficiency benefits of the 55 mph speed limit vary markedly, depending on the value attached to human lives and the costs of time lost by driving at 55 mph rather than 65 mph. One estimate of net efficiency benefits is $2.3 billion, while another is minus $3.4 billion.[42]

[41]See Duncan MacRae Jr. and James Wilde, *Policy Analysis for Public Decisions* (North Scituate, MA: Duxbury Press, 1979), pp. 133–152; Edward R. Tufte, *Data Analysis for Politics and Policy* (Englewood Cliffs, NJ: Prentice Hall, 1974), pp. 5–18; Charles T. Clotfelter and John C. Hahn, "Assessing the National 55 mph Speed Limit," *Policy Sciences* 9 (1978): 281–294; and Charles A. Lave and Lester B. Lave, "Barriers to Increasing Highway Safety," in *Challenging the Old Order: Toward New Directions in Traffic Safety Theory*, ed. J. Peter Rothe (New Brunswick, NJ: Transaction Books, 1990), pp. 77–94. The author has tested the effects of unemployment and gasoline prices on mileage and fatality rates, explaining over 90 percent of the variance in mileage death rates.

[42]See Grover Starling, *Strategies for Policy Making* (Chicago, IL: Dorsey Press, 1988), pp. 407–410, and Guess and Farnham, *Cases in Public Policy Analysis*, pp. 177–203.

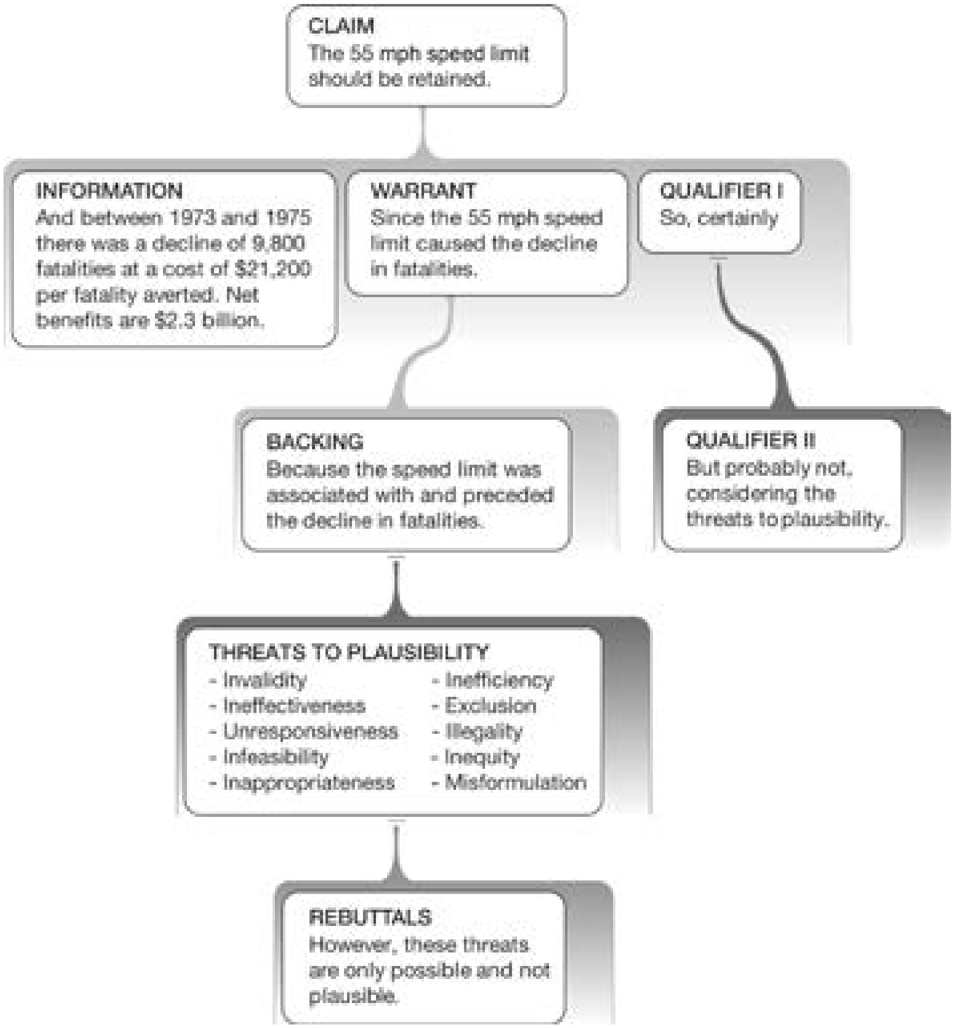


FIGURE 5.11
Threats to the Plausibility of Claims about the Benefits of the 55 mph Speed Limit

- *Ineffectiveness.* Estimates of the cost–effectiveness of the 55 mph speed limit also vary markedly. The costs per fatality averted range from approximately $1,300 to $21,000, depending on assumptions about what should be included as measures of the value of a human life.[43]
- *Exclusion.* The exclusion of legitimate costs and benefits will produce implausibly high or low net efficiency benefits. For example, costs of time lost by driving at

[43]Starling, *Strategies for Policy Making*, pp. 409–411; see also footnote 41.

55 mph or benefits in the form of the monetary value of human lives may be excluded, producing implausibly low net efficiency benefits.

- *Unresponsiveness.* The costs of time and other resources are often based on assumptions that individuals would be willing to pay the average wage rate to engage in a time-saving activity. The average wage rate is typically unresponsive to the price individuals are actually willing to pay. For example, approximately half of drivers are engaged in recreational activities, not commerce or business. The average wage rate, which markedly increases the estimated costs of the 55 mph speed limit, is therefore unresponsive to the costs of driving at 55 mph, as perceived by drivers.
- *Illegality.* In some cases, the legality of a policy may be challenged. It is not appropriate to count as costs (or benefits) income earned through illegality, fraud, or unlawful discrimination. In cases involving the comparable worth of jobs held by men, women, and minorities, illegality is a major factor. In the case of the 55 mph speed limit, it is not.
- *Unfeasibility.* Policy prescriptions can be challenged on grounds that they are not feasible because of political, budgetary, and administrative constraints. The implementation of the 55 mph speed limit varies across states because of differing capacities for implementing or enforcing speed laws. Implementation capacity is generally greater in eastern states with high population density.
- *Inequity.* Prescriptions may be challenged on ethical grounds involving alternative conceptions of social equity and justice. The group that pays a disproportionately high cost in terms of lives lost is younger drivers. The median age of traffic fatalities is slightly more than 30 years, which means that half of traffic deaths are below 30 years of age. The Rawlsian conception of justice as fairness—that is, those worst off should benefit—may be used to challenge policies designed to reduce fatalities through speeding laws rather than through educational and driver training programs that target younger drivers.[44]
- *Inappropriateness.* Prescriptions based on estimates of the value of human lives sometimes employ discounted lifetime earnings as a measure of value. Efforts to establish the cost (discounted or undiscounted) of a human life may be challenged on grounds that it is inappropriate to calculate a price for human lives, which are not commodities on an open market.[45]
- *Misformulation.* A standing challenge to prescriptions based on cost–benefit analysis is that the problem has been misformulated. In the case of the 55 mph speed limit, saving fuel during the oil crisis of 1970–1973 was the original problem for which the National Maximum Speed Law was a prescribed solution. The definition of the problem shifted to averting fatalities after the oil crisis passed. The existing formulation of the problem (averting traffic fatalities) may be challenged on grounds that the problem should be formulated as one of saving petroleum, a nonrenewable resource; mitigating the emission of pollutants; and averting fatalities by increasing gasoline taxes to reduce driving. In inflation-adjusted dollars, the cost of a gallon of gasoline in 2015 ($2.36) was actually lower than during 1931 and the great depression

---

[44]On this question, see, especially, Peter Rothe, ed., *Challenging the Old Order: Toward New Directions in Traffic Safety Theory* (New Brunswick, NJ: Transaction Books, 1990).

[45]See the discussion in Guess and Farnham, *Cases in Public Policy Analysis*.

($2.65) and during 2009 and the great recession ($2.54). Considering that inflation-adjusted per capita income nearly tripled between 1960 and 2015, an increase in gasoline taxes may be a solution.

These threats to plausibility do not apply solely to CBA. Because all policy prescriptions are based on causal as well as value premises, these threats to plausibility are relevant to almost any policy that seeks to achieve change through the regulation, allocation, or reallocation of scarce resources.[46]

The methodology of prescription involves many uncertainties. One of these has to do with the role of values and ethics in policy analysis. The purpose of a policy prescription is not simply to forecast or predict some future outcome or to monitor its past or present effects—but to advocate a course of action the likely consequences of which are *valuable* because they are right in and of themselves or because they promote justice, equality, efficiency, democracy, liberty, security, or another end. However, there are practically significant difficulties in using economic theory and the tools of economic analysis to justify such claims about what is valuable for members of a community. For this reason, uncertainties about values are fit subjects for reasoned ethical argument and debate.

A closely related source of uncertainty stems from incomplete knowledge about the effects of policies. Even if there were a consensus on important social, economic, and political values, we still may not know which policies work best under different conditions. Some of this uncertainty is a result of incomplete information about the range of costs and benefits that should be considered in a systematic analysis. Another important source of uncertainty stems from measures of varying reliability and validity and fallible judgments about shadow prices and discount rates.

CHAPTER SUMMARY

Policy prescriptions answer the question: What should be done? For this reason, policy prescriptions require an approach that is normative, and not one that is merely descriptive. All policy prescriptions involve claims about right action, rather than claims that are simply factual in character.

This chapter has provided an overview of the methodology of prescription in policy analysis, compared and contrasted two major approaches to prescription, and described methods used to carry out these approaches. Two major approaches to prescription in policy analysis are cost–benefit analysis and cost–effectiveness analysis. While both approaches seek to measure costs and benefits to society, only cost–benefit analysis measures benefits and costs in a common unit of monetary value.

Costs and benefits are of several types: internal versus external, tangible versus intangible, primary versus secondary, and real versus pecuniary. There are several tasks in cost–benefit analysis: problem structuring; specification of objectives; identification of alternatives; collection, analysis, and interpretation of information; specification of target groups and beneficiaries; identification of types of costs and benefits; discounting of costs and benefits; and specification of criteria for choosing among policies. The criterion most

[46]See Dunn, "Policies as Arguments"; and Frank Fischer, *Politics, Values, and Public Policy: The Problem of Methodology* (Boulder, CO: Westview Press, 1980).

frequently employed in traditional cost–benefit analysis is net efficiency improvement. Contemporary cost–benefit analysis supplements this criterion with that of net redistributional benefits.

REVIEW QUESTIONS

1. The late Milton Friedman, Nobel Laureate in Economics from the University of Chicago, argued for a *positive* (rather than *normative*) economics that focuses on the study of the factual premises underlying policy choices:

   "Differences about economic policy among disinterested citizens derive predominantly from different predictions about the economic consequences of taking action—differences that in principle can be eliminated by the progress of positive economics—rather than from fundamental differences in basic values, differences about which men can only fight."
   (Friedman [1953], p. 5)

   Write a short essay on the strengths and weaknesses of Friedman's position. In your answer refer to contrasts between descriptive and normative analysis (Chapter 1). [Note: If you wish to read ahead (Chapter 7), assess Friedman's position in terms of contrasts among descriptive ethics, normative ethics, and meta-ethics].

2. "If individuals can order their preferences in a transitive fashion, it should be possible to use majority rule to obtain a collective preference ranking which is transitive." Discuss.
3. Is rational choice possible? Write an essay on this question. In your answer distinguish different types of rationality.
4. How are different types of rationality related to different criteria for prescription?
5. Indicate how you would measure the effectiveness of efforts to achieve the following goals:
   - Close the gap between municipal revenues and expenditures
   - Deter the commission of serious crimes in urban areas
   - Enhance the national security of the country
   - Win the war on terror(ism)
   - Improve the quality of life of citizens
   - Reduce national unemployment
   - Increase the income of poverty families
6. Return to Question 5. Indicate how you would measure the *efficiency* of efforts to achieve these goals.
7. Many economists argue that equality and efficiency are competing objectives. For this reason, the relation between equality and efficiency is seen as a tradeoff, where gains in equality are sacrificed for gains in efficiency, and vice versa [see Arthur M. Okun, *Equality and Efficiency: The Big Trade-Off* (Washington, DC: Brookings Institution, 1975)]. Other economists [see Jaroslav Vanek, *The Participatory Economy* (Ithaca, NY: Cornell University Press, 1970)] and political scientists [see Carole Pateman, *Participation and Democratic Theory* (Cambridge, MA: Cambridge University Press, 1970)] argue that the "tradeoff" between equality and efficiency is a consequence of the limited forms of democratic participation in politics and work. Equality and efficiency, they argue, do not have to be "traded" when politics and economics are fully democratic. Why would this be true? What does it suggest about the appropriateness of economic concepts such as "opportunity costs" and "net efficiency benefits?"
8. Return to Figure 5.1 and consider the following criteria for prescription:
   a. Maximize effectiveness at least cost [*Note:* Be careful—this is a tricky question].

b. Maximize effectiveness at a fixed cost of $10,000.
c. Minimize costs at a fixed-effectiveness level of 4,000 units of service.
d. Achieve a fixed-effectiveness level of 6,000 units of service at a fixed cost of $20,000.
e. Assuming that each unit of service has a market price of $10, maximize net benefits.
f. Again assuming that each unit of service has a market price of $10, maximize the ratio of benefits to costs.

Indicate which of the two main programs (Program I and Program II) should be selected under each of these criteria, and describe the conditions under which each criterion may be an adequate measure of the achievement of objectives.

9. The estimation of risk and uncertainty is an essential aspect of prescription. While it seems obvious that the best way to obtain valid estimates of uncertainty is to use the most highly qualified experts, there is evidence that "people who know the most about various topics are not consistently the best at expressing the likelihood that they are correct" [Baruch Fischoff, "Cost-Benefit Analysis and the Art of Motorcycle Maintenance," *Policy Sciences* 8 (1977): p. 184]. Discuss.
10. Use procedures of values clarification and values critique to analyze value premises that underlie the following claim: "Increased energy production is essential to the continued growth of the economy. The government should therefore make massive investments in research and development on new energy technologies." Treat the first sentence as information (*I*) and the second as the claim (*C*).
12. Calculate the discount factor (DF) at 10 years for a 3, 4, and 10 percent discount rate.
13. Calculate the present value (PV) of the following cost and benefit streams (in millions of dollars) for two health programs. Use a discount rate of 4 percent. Which program has the highest present net efficiency benefits? Why?

	Year					
Program	*2000*	*2005*	*2010*	*2015*	*2020*	*Total*
I Costs	$60	$10	$10	$10	$10	$100
Benefits	10	20	30	40	50	150
II Costs	20	20	20	20	20	100
Benefits	10	20	30	40	50	150

DEMONSTRATION EXERCISES

1. After reading Case 5.1, compare, contrast, and evaluate the scorecard and the spreadsheet as forms of analysis. What are their strengths and weaknesses? Be specific.
2. After reading Case 5.2 (Opportunity Costs of Saving Lives—The National Maximum Speed Limit), revise the spreadsheet (Table C.5.1) after answering these questions.
   - What assumptions govern estimates of the value of time lost driving? Are some assumptions more tenable than others? Why?

- What is the best way to estimate the value of time? Justify your answer.
- What is the best way to estimate the cost of a gallon of gasoline? Justify your answer.
- Driving speeds and miles per gallon estimates are based on official statistics from the Environmental Protection Agency and the Department of Energy, or on engineering studies of the efficiency of gasoline engines. Which is the more reliable? Why? What are the consequences of using one source rather than another?
- What is the value of a life saved? How do you calculate that value? Why?
- Which policy is preferable, the 55 mph speed limit, or the 65 mph limit? Justify your answer.

3. Case 5.2 shows some of the hazards of conducting a cost–benefit analysis without considering the changes in measurable outcomes over time. Contrast the use of time in Case 5.3 with the use of time when discounting future costs and benefits. Could you estimate the preset value of cost and benefit streams over the period 1974–2015? What obstacles would you need to overcome?

BIBLIOGRAPHY

Boardman, Anthony E., David H. Greenberg, Aidan R., Vining, and David L. Weimer. *Cost-Benefit Analysis: Concepts and Practice*. 4th ed. Upper Saddle River, NJ: Pearson-Prentice Hall, 2011.

Burt, M. R. *Policy Analysis: Introduction and Applications to Health Programs.* Washington, DC: Information Resources Press, 1974.

Clotfelter, Charles T. and John C. Hahn. "Assessing the National 55 m.p.h. Speed Limit." *Policy Sciences* (1978): 281–294.

Fischhoff, Baruch. "Cost-Benefit Analysis and the Art of Motorcycle Maintenance," *Policy Sciences* 8 (1977): 177–202.

Friedman, Milton. *Essays in Positive Economics.* Chicago, IL: University of Chicago Press, 1953.

George, Alexander. *Presidential Decision Making in Foreign Policy: The Effective Use of Information and Advice*. Boulder, CO: Westview Press, 1980.

Gramlich, Edward M. *Benefit-Cost Analysis of Public Programs*, 2nd ed. Englewood Cliffs, NJ: Prentice Hall, 1990.

Haveman, Robert H., and Julius Margolis, eds. *Public Expenditure and Policy Analysis*, 3rd ed. Boston: Houghton Mifflin, 1988.

Hinrichs, H. H., and G. M. Taylor. *Systematic Analysis: A Primer on Benefit-Cost and Program Evaluation*. Pacific Palisades, CA: Goodyear Publishing, 1972.

Lave, Charles A. "The Costs of Going 55." *Car and Driver*. May 1978, p. 12.

Miller, Roger Le Roy, Benjamin, Daniel K., and Douglass C. North. *The Economics of Public Issues*, 16th ed. New York: Addison-Wesley, 2010.

Mishan, Edward J. *Cost-Benefit Analysis*. New York: Praeger, 1976.

Moore, Mark A., Anthony E. Boardman, Aidan R. Vining, David L. Weimer, and David H. Greenberg. "'Just Give Me a Number!' Practical Values for the Social Discount Rate." *Journal of Policy Analysis and Management* 23, 4 (2004): 789–812.

Piketty, Thomas. *Capital in the 21st Century*. Cambridge, MA: Harvard University Press, 2014.

Quade, E. S. *Analysis for Public Decisions*. New York: American Elsevier, 1975.

Zeckhauser, Richard, and Elmer Schaefer. "Public Policy and Normative Economic Theory," in *The Study of Policy Formation,* ed. R. A. Bauer and K. J. Gergen. New York: The Free Press, 1968.

CASE 5.1 THE SPREADSHEET AND SCORECARD—EVALUATING BENEFITS AND COSTS OF ENERGY POLICIES

In 1972 and 1973, the United States and other petroleum-dependent countries in Europe experienced the first of several oil crises precipitated by a dramatic increase in the price of crude oil by the Organization of Petroleum Exporting Countries (OPEC). One response of American and European leaders was to adopt maximum speed limits of 55 mph and 90 kph, respectively. In the United States, the National Maximum Speed Law (NMSL) of January 1, 1974 was designed to reduce the consumption of gasoline by requiring that all vehicles on interstate highways travel at a maximum speed of 55 mph, a speed which at that time maximized fuel efficiency for most vehicles.

TABLE C5.1

Scorecard and Spreadsheet

(a) Scorecard

OUTCOMES	65mph (Base Case)*	55mph
Fatalities	54,052	**45,196**
Miles traveled (billions)	1,313	**1,281**
Hours driving (billions)	**20.2**	21.9
Gallons fuel consumed (billions)	46.8	**43.3**
Fuel efficiency (mpg)	14.9	**16.1**
Traffic citations (millions)	5,711	**7,425**
Property damage (million cases)	25.8	**23.1**

*The base case is the policy against which the new policy is compared.

(b) Spreadsheet

OBJECTIVES	65mph	55mph	Difference	Value	$ Billions
I Fatalities (000s)	54.1	45.2	8.856	$240,000.00	$ 2.13
II Hours driving (billions)	20.2	21.9	−1.7	5.05	−8.59
III Gallons fuel consumed (billions)	46.8	43.3	3.5	0.53	1.86
IV Traffic citations (000s)	5,711	7,425	−1,714	3.94	−0.0068
V Property damage cases (000s)	25,800	23,100	2,700	363.00	0.98
Benefits (I + III + V)					4.97
Costs (II + IV)					−8.59
Net Benefits (B−C)					$ −3.63

Soon after the implementation of the 55 mph speed limit it was discovered that the new policy not only reduced fuel consumption, but apparently caused a dramatic decline in traffic fatalities and injuries. Long after the OPEC oil crisis was over, the speed limit was retained, although it was no longer needed to respond to the energy crisis that prompted its passage in 1973. Indeed, the 55 mph speed limit, created in response to the energy crisis of 1972–1975, was retained for more than 20 years until it was officially repealed in November 1995.[47]

Heated debates preceded the repeal. Senator John C. Danforth of Missouri, an influential advocate of the policy, argued that the repeal would save one minute per day per driver but result in an additional 600 to 1,000 deaths. *The Washington Post* and *The New York Times* joined the opposition, reporting that, while the savings in time were trivial, fatalities would surely rise. Later, Secretary of Transportation Federico Pena announced that the Clinton Administration was firmly opposed to abandoning the speed limit.

This was the right moment for an evaluation of the benefits and costs of the National Maximum Speed Limit. The scorecard is a useful tool for monitoring and forecasting impacts under conditions where benefit–cost analysis is not feasible or desirable. In the scorecard, policy alternatives are arrayed in columns along the top of the matrix and policy impacts are listed in each row. Spreadsheets, by contrast, display the benefits and costs of policy outcomes.

Table C5.1(b) displays a spreadsheet used to evaluate the effects of the 55 mph speed limit at the end of 1974, 1 year after the policy was implemented. To show the differences between the spreadsheeet and the scorecard, Table C5.1 (a) also displays the same information as a scorecard.

Note that the highlighted cells of the scorecard (C5.1 (a)) show that the 55 mph speed limit is more advantageous than the 65 mph speed limit (5 of 7 rows are highlighted). However, the spreadsheet (b) tells a different story. Using a cost–benefit framework, the spreadsheet shows that the 55 mph speed limit has *negative* net benefits. The 65 mph speed limit is more efficient. ■

CASE 5.2 OPPORTUNITY COSTS OF SAVING LIVES—THE 55 MPH SPEED LIMIT[48]

Conducting a benefit–cost analysis is more than a technical exercise accomplished with a spreadsheet. It is also a matter of identifying, and if necessary challenging, the assumptions on which a benefit–cost analysis is based. This can be seen when we examine the case of the National Maximum Speed Limit of 1974.

Table C5.2 describes steps in conducting a benefit–cost analysis of the 55 mph speed limit and a critique of the assumptions underlying the analysis. The case shows, among other things, that the final step in conducting a benefit–cost analysis—namely, the calculation of benefit–cost ratios and net efficiency benefits—is quite sensitive to these assumptions. This kind of critical analysis is a prerequisite for preparing the data for the spreadsheet presented in Case 5.1. ■

[47]On April 2, 1987, Congress enacted the Surface Transportation and Uniform Relocation Assistance Act, permitting forty states to experiment with speed limits up to 65 mph.

[48]Adapted from Grover Starling, *Strategies for Policy Making* (Chicago, IL: Dorsey, 1988), pp. 407–410, and William N. Dunn, "A Pragmatic Strategy for Discovering and Testing Threats to the Validity of Sociotechnical Experiments." *Simulation Modelling Practice and Theory* 10, 3 (2002): 169–194.

TABLE C5.2

Analysis of the Costs and Benefits of the 55 mph Speed Limit

Steps	Critique
Estimate Costs	
1. The major cost of the National Maximum Speed Law (NMSL) was the additional time spent driving as a result of slower speeds. To calculate the number of hours spent driving in 1973, the year before the 55mph speed limit, divide the total number of vehicle miles driven (1,313 billion) by the average highway speed (65 mph) and then multiply by the average occupancy rate of approximately 1.5 per vehicle.	
Next, find the number of hours spent driving in 1974, the year the NMSL was implemented, by dividing total vehicle miles (1,281 billion) by the average highway speed in 1974 (58 mph). Because the NMSL caused some people to cancel trips and others to find alternative modes of transportation, time calculations based on 1974 mileage are an underestimate. Therefore, we should use time calculations based on 1973 mileage.	
Using the following formula, where *VM* is vehicle miles, *S* is average speed, *R* is average occupancy rate, and *H* is the number of hours lost at the lower speed:	*Why use 1973 mileage without any adjustment? The average growth rate in travel before 1973 was 4 percent. Hence, the formula should be:*
$H = \left(\frac{VM_{1973}}{S_{1974}} - \frac{VM_{1973}}{S_{1973}} \right) \times R$	$H = \left(\frac{1.04 VM_{1973}}{S_{1974}} - \frac{VM_{1973}}{S_{1973}} \right) \times R$
The number of hours lost driving in 1974 is estimated to be **3.6 billion** hours.	*Using the above formula, the estimated number of hours lost should be* **5.02 billion**, *not 3.6 billion.*
2. To estimate the value of this time, begin with the average wage rate for all members of the labor force in 1974—$5.05. The value of one hour's travel is not $5.05 per hour because very few persons would pay this sum to avoid an hour of travel. We estimate that the people will pay up to 33 percent of their	*Why take a percentage of the $5.05 figure based on what commuters would pay to avoid an hour of travel? We should avoid reducing the value of people's time for two reasons. First, the value of time in cost to society is equal to what society will pay for productive use of that time. Time's value is not what a commuter will pay to avoid commuting because*

(*continued*)

average hourly wage rate to avoid an hour of commuting. The value of time spent traveling is therefore 0.33 × \$5.05 = \$1.68 per hour.	*commuting has other benefits, such as solitude for thinking or the advantages of suburban living. Second, the value of time spent driving for a trucker is many times the industrial wage rate. Discounting would greatly underestimate the value of commercial drivers.*
3. Application of the cost figure (\$1.68) to the time lost figure (1.72 billion hours) results in an estimated travel cost of ***\$2.89 billion***.	*Applying the value of one hour's time (\$5.05) to the hours lost as calculated above (5.02 billion) results in an estimated travel cost of \$25.35 billion.*
4. The NMSL also has some enforcement costs. Total enforcement costs for signs, advertising, and patrolling are about \$810,000.	*Studies indicate that total enforcement costs should be about* **\$12 million***—not \$810,000.*
a. New signs were posted. Cost estimates from twenty-five states for modification of speed limit signs totaled \$707,000; for fifty states, this results in an estimated \$1.23 million. Spread out over the 3-year life of traffic signs, we get an estimate of ***\$410,000***.	***OK***.
b. The federal government engaged in an advertising campaign encouraging compliance. The Federal Highway Administration's advertising budget for 1974 was \$2 million. About 10 percent of this, or \$200,000, was spent to encourage compliance with the NMSL. Assume that an additional amount of public service advertising time was donated, for a total of ***\$400,000***.	*Not OK. The Federal Highway Administration does other advertising; not all \$200,000 should be counted part of NMSL. Public service advertising estimate also seems low.*
c. Compliance costs are difficult to estimate. The cost of highway patrols cannot be used because these persons were patrolling highways before the NMSL. Assume that states did not hire additional personnel solely for enforcement of the NMSL. Therefore, we assume that enforcement of the NMSL will not entail any additional costs above enforcement of previous speed limits.	*In 1973, some 5,711,617 traffic citations jumped by 1,713,636 to over 7.4 million. Each additional traffic citation includes an opportunity cost to society. If a law enforcement officer were not issuing traffic tickets, he could be solving other crimes. Assuming that it requires 15 minutes for a law enforcement officer to issue a speeding ticket, the total cost of law enforcement is 1.7 million citations × 0.25H =* **450,000 hours***. The average cost of placing a law enforcement officer on the streets is \$6.75 per hour, so the cost is \$450,000 × \$6.75 =* **\$3.03 billion***. Even this figure is an underestimate because it does not count time lost waiting to catch speeders.*

Approximately 10 percent of all speeders will demand a court hearing. Estimating an average of 30 minutes for each hearing and an hourly court cost of $45 results in an additional cost to society of **$3.8 million** *for 171,000 cases (171,000 × $45.00/2). Given the overloaded court dockets, this opportunity cost may be even higher.*

Estimate Benefits

1. The most apparent benefit of the NMSL is the amount of gasoline saved. Average gasoline economy improves from 14.9 mpg at 65 miles per hour to 16.1 mpg at 58 miles per hour. To estimate the number of gallons of gasoline saved by traveling at lower speeds, use the following formula:

$$G = \frac{VM_{1973}}{MPG_{1973}} - \frac{VM_{1973}}{MPG_{1974}}$$
$$= \frac{1313 \text{ bil. mi.}}{14.9 \text{ mpg}} - \frac{1313 \text{ bil. mi.}}{16.1 \text{ mpg}}$$

Why estimate gasoline saved by comparing 1973 and 1974 miles-per-gallon figures in relation to vehicle miles traveled? The federal figures for average miles per hour are estimates based on several assumptions. Given the conflict between industry estimates, Environmental Protection Agency estimates, and Energy Department estimates, any miles-per-hour estimate must be considered unreliable. The number of vehicle miles traveled is also based on gallons of fuel sold multiplied by average miles per hour. Hence, this figure is also subject to error.

Studies of the efficiency of gasoline engines show that the effect of reducing the average speed of free-flow interstate highways would save 2.57 percent of the normal gas used. In 1979, American motorists consumed 106.3 billion gallons of gasoline. Saving 2.57 percent would total **2.73 billion gallons** *(106.3 billion gals × 0.0257 = 2.73).*

In 1974, the average price of gasoline was 52.8 cents per gallon. This market price, however, does not reflect the social cost of gasoline, due to government price controls on domestic oil. The marginal (or replacement) cost of crude oil is the price of foreign oil. Therefore, the price of gasoline must reflect the higher cost of foreign oil. Use the market price of gasoline in the absence of price controls, which is about 71.8 cents per gallon. This figure yields an estimate of **$2.51 billion** (71.8 cents × 3.5 billion gals) in benefits through gasoline saved.

Why not use the market price? There is no way to determine whether a marginal gallon of gasoline will be imported or come from domestic reserves. In addition, the costs and benefits of the NMSL should not be distorted simply because the U.S. government does not have a market-oriented energy policy.

In 1974, gasoline cost 52.8 cents per gallon, and therefore, a gallon of gasoline saved was worth 52.8 cents. (2.73 billion gals × $0.52 = ***$1.45 billion.)***

(*continued*)

2. A major second-order benefit of the NMSL was a decrease of 9,038 traffic fatalities, from 55,087 in 1973 to 46,049 in 1974. Part of the decrease is attributable to reduction in traffic speeds. Studies by the National Safety Council estimate that up to 59 percent of the decline in fatalities was the result of the speed limit. Applying this proportion to the decline in fatalities provides an estimated 5,332 (9,038 × 0.59) lives saved. The consensus of several studies is that a traffic fatality costs $240,000 in 1974 dollars. Using this figure, the value of lives saved in 1974 is estimated at **$1,279.7 billion** (5,332 × $240,000). — ***OK***

3. The NMSL also resulted in a reduction of nonfatal injuries. Use the 59 percent figure found in the fatality studies. Between 1973 and 1974, nonfatal traffic injuries declined by 182,626. Applying the estimated percentages results in 107,749 injuries avoided. Generally, three levels of injuries are identified: (1) permanent total disability, (2) permanent partial disability and permanent disfigurement, and (3) nonpermanent injury. In 1971, the proportion of traffic injuries that accounted for injuries in each category was .2 percent, 6.5 percent, and 93.3 percent, respectively. The National Highway Traffic Safety Administration estimated that in 1971 the average cost of each type of injury was $260,300, $67,100, and $2,465, respectively. The average injury, therefore, cost $8,745 in 1974 dollars. Applying this figure to our injury estimate results in **$942.3 million** ($8,745 × 107,749) as the social benefit of injury reduction. — ***OK***

4. The final benefit of the reduction in property damage fell from $25.8 million to $23.1 million. About 50 percent of this reduction was the result of lower speeds. The NMSL saved 1.3 million cases of property damage at an average cost of $363. Therefore, the total benefits from property damage prevented is ***$472 million* (1.3 million × $363).** — ***OK***

Conclusion

One estimate of the costs and benefits of the NMSL is as follows (all figures in billions):

Costs	
Time spent traveling	$2.9
Enforcement	$0.000810
	$2.9
Benefits	
Gasoline saved	$2.51
Lives saved	$1.280
Injuries prevented	.000942
Property damage averted	.000472
	$6.69

Net benefits: $6.69 – 2.9 = $3.79 billion

Benefits to costs ratio: 2.3

A rival estimate of costs and benefits of the NMSL is as follows (all figures in billions):

Costs	
Time spent traveling	$9.8
Enforcement	$0.000012
	$9.8
Benefits	
Gasoline saved	$1.45
Lives saved	$1.28
Injuries prevented	$.000942
Property damage	$.000472
	$2.74

Net benefits: $2.74 – 9.8 = –$7.06

Benefits to costs ratio: 0.28

Source: The steps and data were suggested by Clotfelter and Hahn (1978): 281–294. The critical comments are based on Lave (1978): 12.

CHAPTER

6

Monitoring Observed Policy Outcomes

LEARNING OBJECTIVES

By studying this chapter, you should be able to:

- Distinguish monitoring from other policy-analytic methods
- Describe the main functions of monitoring
- Distinguish policy inputs, activities, outputs, outcomes, and impacts
- Compare and contrast social systems accounting, social auditing, policy experimentation, research and practice synthesis, systematic review, and meta-analysis
- Define and illustrate threats to the validity of experiments, quasi-experiments, and natural experiments
- Perform interrupted time-series and control-series analyses to explain policy outcomes

INTRODUCTION

The consequences of policies are never fully known in advance. For this reason, the policy-analytic procedure of monitoring is essential to policy analysis. Indeed, much of the work of policy analysis is carried out after policies have been prescribed and adopted.[1] Although prescription provides valuable information about what should be done to solve a problem, only by monitoring the outcomes of a policy can we find out whether the policy has been successfully implemented.[2] Monitoring helps answer a profoundly important practical question: Does the policy work?[3]

MONITORING IN POLICY ANALYSIS

Monitoring permits the production of information about the causes and consequences of policies. Because monitoring investigates relations between policy operations and their observed outcomes, it is the primary source of information about the success of efforts to implement policies. Monitoring goes beyond the normative prescriptions of economic analysis by establishing factual premises about policies. Monitoring is primarily concerned with establishing factual premises about public policies, while evaluation (Chapter 7) is concerned with establishing value premises. While factual and value premises are in continuous flux, and "facts" and "values" are interdependent, only monitoring produces factual claims during and after policies have been adopted and implemented (*ex post*).

Monitoring performs several functions in policy analysis: compliance, auditing, accounting, and description and explanation.

- *Compliance.* Monitoring helps determine whether the actions of program managers are in compliance with norms, values, and standards mandated by legislatures, regulatory agencies, and professional associations. For example, the Environmental Protection Agency's Continuous Air Monitoring Program (CAMP) produces information about pollution levels to determine whether industries are complying with federal air quality standards.
- *Auditing.* Monitoring helps discover whether resources and services intended for target groups and beneficiaries actually reach them, for example, recipients of social services or municipalities that qualify for federal and state grants. By monitoring federal revenue sharing, we can determine the extent to which funds are reaching local governments.[4]
- *Accounting.* Monitoring produces information that is helpful in accounting for social and economic changes that follow the implementation of policies over

[1]Graham T. Allison, *Essence of Decision: Explaining the Cuban Missile Crisis* (Boston, MA: Little, Brown, 1971).
[2]A unique source of studies of evidence-based policies that "work" is the Campbell Collaboration (www.campbellcollaboration.org) and the Cochrane Collaboration in medicine and health policy (www.cochranecollaboration.org).
[3]A useful practical framework for implementation analysis is R. Kent Weaver, "But Will It Work? Implementation Analysis to Improve Government Performance," *Issues in Governance Studies* 32 (February 2010), pp. 1–17.
[4]Richard P. Nathan and others, *Monitoring Revenue Sharing* (Washington, DC: Brookings Institution, 1975).

time. For example, changes in well-being may be monitored with such social indicators as average education, percentage of the population below the poverty line, and average annual paid vacations.[5] The World Bank, the European Union, the Organization of African Unity, the United Nations, and the United States, and other national and international bodies employ social indicators to monitor changes in quality of life.[6]

- *Description and explanation.* Monitoring also yields information that helps to explain why the outcomes of public policies and programs result in particular outcomes. The Campbell Collaboration, which has spread to more than forty countries, is an archive of field experiments in education, labor, criminal justice, housing, social work, and other areas. The Campbell Collaboration and the parallel organization in medicine and public health, the Cochrane Collaboration, aim at evidence-based policy and evidence-based medicine.[7]

Sources of Information

To monitor policies we require information that is relevant, reliable, and valid. If we want to know about the consequences of programs designed to provide educational opportunity for disadvantaged children, we need information that not only documents obstacles to achieving educational opportunity in general, but information that yields answers about specific factors that policymakers may manipulate to induce greater opportunity. The first type of information is appropriately termed *macronegative*, while the latter is best described as *micropositive*. Information acquired through monitoring must also be reliable, which means that observations should be reasonably precise, dependable, and internally consistent. For example, we have known for many years that information on crime is unreliable by a factor of about 2.5 to 1; that is, there are some two to three times more crimes actually committed than reported to police.[8] Finally, we also want to know whether information about policy outcomes is actually measuring what we think it is, that is, whether the measurements are valid. If we are interested in violent crimes, for

[5]See, for example, Angus Deaton, "Income, Health, and Well-Being Around the World: Evidence from the Gallup World Poll." *The Journal of Economic Perspectives* 22, 2 (2008): 53–72; and Tom Rath and Jim Harter, *Wellbeing: The Five Essential Elements* (New York: Simon and Schuster, 2010).

[6]See World Bank, *World Development Indicators* at http://data.worldbank.org/indicator; Anthony Barnes Atkinson, *Social Indicators: The EU and Social Inclusion* (New York: Oxford University Press, 2003); Duncan MacRae Jr., *Policy Indicators: Links between Social Science and Public Debate* (Chapel Hill: The University of North Carolina Press, 1985); Office of Management and the Budget, *Social Indicators, 1973* (Washington, DC: U.S. Government Printing Office, 1974).

[7]The preeminent source on these questions is Will Shadish, Thomas D. Cook, and Donald T. Campbell, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference* (Belmont, CA: Wadsworth, Cengage Learning, 2002). See also, William N. Dunn, ed. and contributor, *The Experimenting Society: Essays in Honor of Donald T. Campbell* (New Brunswick, NJ: Transaction, 1998). Early sources in this tradition are Alice Rivlin, *Systematic Thinking for Social Action* (Washington, DC: Brookings Institution, 1971); and George W. Fairweather and Louis G. Tornatzky, *Experimental Methods for Social Policy Research* (Oxford and New York: Pergamon Press, 1977).

[8]The original estimates are in R. H. Beattie, "Criminal Statistics in the United States," *Journal of Criminal Law, Criminology, and Police Science* 51 (May 1960): 49–51.

example, information on crimes in general (which includes auto theft and white-collar crimes) will not be a valid measure of the kinds of policy outcomes in which we are interested.[9] Other prominent monitoring systems are the Worldwide Governance Indicators (WGI). The validity of the WGI, defined in terms of convergent and discriminant validity, has been investigated with results suggesting that this monitoring system has limited validity.[10]

Information on policy outcomes is regularly collected at various points in time at considerable cost to federal, state, and local governments, private research institutes, and universities. Some of this information is general—for example, information about the social, economic, and demographic characteristics of the population as a whole—and some is more specific because it deals with regions, states, municipalities, and other subpopulations within society. Information about policy outcomes is published by the U.S. Bureau of the Census, the U.S. Bureau of Labor Statistics, offices and clearinghouses attached to federal agencies and institutes, several nonprofit research centers, and agencies of the European Union and the United Nations.[11]

In addition to these sources, federal, state, and municipal agencies produce reports on special programs and projects in education, health, welfare, labor, employment, crime, energy, pollution, foreign affairs, and other areas. Research centers, institutes, and universities also maintain data archives that contain a wide variety of political, social, and economic data, and there is a large stock of books, monographs, articles, and reports written by applied researchers throughout the country. Much of this information can now be accessed through many online information systems, including *JSTOR* for academic publications and *PolicyFile,* a unique resource for U.S. public policy research. *PolicyFile* contains timely, updated information from over 350 public policy think tanks, nongovernmental organizations, research institutes, university centers, advocacy groups, and other entities. Over seventy-five public policy topics are covered, from foreign policy to domestic policy.[12] *Google Scholar* also covers peer-reviewed academic studies of policies and programs as well as studies produced by government agencies at the national and international levels and nongovernmental organizations of many kinds.

When data and other types of information are not available from existing sources, monitoring may be carried out by some combination of questionnaires, interviews, field observations, and the use of agency records.[13] Because most analysts do not conduct original research—given that the funding and time required for such research are

---

[9]See Wesley G. Skogan, "The Validity of Official Crime Statistics: An Empirical Investigation," *Social Science Quarterly* 55 (June 1974): 25–38.

[10]Melissa A. Thomas, "What Do the Worldwide Governance Indicators Measure?" *The European Journal of Development Research* 22, 1 (2010): 31–54.

[11]For guides to data references, archives, social indicators, and scales and indexes, see Delbert C. Miller and Neil J. Salkind, *Handbook of Research Design and Social Measurement*. 6th ed. (Beverley Hills, CA: Sage Publications, 2002).

[12]The URL for *PolicyFile* is www.policyfile.com.

[13]A thorough and systematic guide to mail, telephone, and electronic surveys is Don A. Dillman, Jolene D. Smyth, and Leah Melani Christian. *Internet, Mail, and Mix-Mode Surveys: The Tailored Design Method* (Hoboken, NJ: John Wiley, 2009).

scarce—existing data archives and published and unpublished reports are some of the major sources for many analyses. When funding for monitoring is available, books and manuals on what we are here calling monitoring fall within the broad area of program evaluation.[14]

Policy-Relevant Causes and Effects

In monitoring policy outcomes there are several kinds of policy-relevant effects. *Policy outcomes* are the goods, services, and resources actually received by target groups and beneficiaries as an effect of the policy or program. However, per capita welfare expenditures and units of packaged food produced by Meals on Wheels are examples of *policy outputs* rather than *policy outcomes*. In contrast to policy outcomes, *policy impacts* are longer-term changes in knowledge, attitudes, or behaviors that result from policy *outcomes*. For example, while units of packaged food consumed by the elderly is an appropriate outcome indicator, the longer-term health and well-being of elderly persons who have consumed the food provided by Meals on Wheels is a measurable *policy impact*. Similarly, the number of hospital beds per 1,000 persons is a *policy output* that is a consequence of numerous human and material *inputs* and *activities*. To monitor policy outcomes and impacts, however, we have to determine how many beneficiaries of hospital admission experienced increases in short and long term health status.

To account for variations in policy outputs, outcomes, and impacts, it is necessary to trace them back to prior policy inputs and activities. The relation between inputs, activities, outputs, outcomes, and impacts may be visualized as a complex causal mechanism (Figure 6.1).[15] Policy *inputs* are the human and material resources—time, money, personnel, supplies, technology—used to produce activities, outputs, outcomes, and impacts. A good example of policy inputs is the program budget, which is a systematic set of accounts that allocate the resources needed to carry out program activities. Policy *activities,* in turn, are the functions of planning, administration, research and development, budgeting, human resources, and training performed in order to produce outputs, outcomes, and impacts. The entire causal mechanism is illustrated in Figure 6.1. Although the figure shows a linear relation, in practice, there are feedback and feedforward loops and short circuits.

In monitoring elements of the causal mechanism, it is important to distinguish between target groups and beneficiaries. *Target groups* are individuals, organizations, communities, regions, or states and provinces on whom policies and programs are designed to have an

---

[14]Representative books and manuals are Peter H. Rossi, Mark W. Lipsey, and Howard E. Freeman, *Evaluation: A Systematic Approach* (Thousand Oaks, CA: Sage Publications, 2003); Carol H. Weiss, *Evaluation*, 2nd ed. (Upper Saddle River, NJ: Prentice Hall, 1998); Michael Quinn Patton, *Essentials of Utilization-Focused Evaluation* (Sage Publications, 2011); and P. J. Gertler, S. Martinez, P. Premand, L. B., Rawlings, L. B., and C. M. Vermeersch, *Impact Evaluation in Practice* (Washington, DC: World Bank Publications, 2011). As noted, the preeminent source on program evaluation, and applied social science generally, is Shadish, Cook, and Campbell, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference* (2002).

[15]The causal mechanism presented in Figure 6.1 is based on descriptive, not explanatory causality, as are most field experiments. A comprehensive discussion of causal mechanisms is provided by John Gerring, "Causal Mechanisms: Yes, But. . . . " *Comparative Political Studies* 43, 1 (2010): 1499–1526.

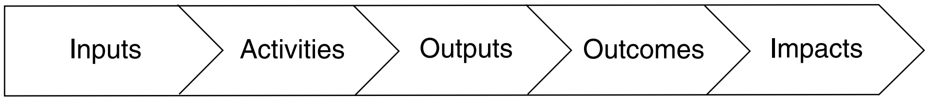


FIGURE 6.1
Causal Mechanism: Inputs, Activities, Outputs, Outcomes, and Impacts

effect. However, *beneficiaries* are groups for whom the effects are beneficial or valuable.[16] For example, industrial and manufacturing firms are the targets of federal programs administered by the Occupational Safety and Health Administration (OSHA), but workers are the beneficiaries.

Policies and programs may be characterized in terms of the organizational and political activities and attitudes that shape the transformation of policy inputs into policy outcomes and impacts. For example, conflicts among agency staff and management, dissatisfaction with working conditions, or inflexible decision-making procedures may explain why programs that have the same resource inputs produce lower levels of outputs. The important point is to distinguish between inputs, activities, and outputs, on the one hand, and outputs and impacts on the other. To fail to do so is akin to measuring "the number of times a bird flaps its wings without any attempt to determine how far the bird has flown."[17] In Table 6.1, inputs, activities, outputs, outcomes, and impacts are illustrated in areas of criminal justice, municipal government, and social welfare.

Definitions and Indicators

Our success in obtaining, analyzing, and interpreting data on policy outcomes depends on our capacity to construct reliable and valid measures. One way to construct measures is to begin by specifying the constructs and variables we are interested in monitoring. A *construct* is a concept that is defined for purposes of social and policy inquiry. Constructs are defined constitutively, by specifying the formal properties that distinguish it from other concepts. For example, returning to the causal mechanism described in Figure 6.1, a construct related to policy inputs is "budget," which may be defined constitutively as *a proposed or formally appropriated set of revenues and expenditures developed in order to support program or policy activities*. This constitutive definition, sometimes called a lexical or "dictionary" definition, precedes the operational definition by stipulating the key properties of a construct.

[16]The distinction between target groups and beneficiaries expresses a contrast similar to that of Lowi, who distinguished between distributive, regulative, and redistributive policies. Theodore J. Lowi, "American Business, Public Policy Case Studies, and Political Theory," *World Politics* 16 (July 1964): 689–690; and Charles O. Jones, *An Introduction to the Study of Public Policy,* 2nd ed. (North Scituate, MA: Duxbury Press, 1977), pp. 223–225. A number of writers conflate target groups and beneficiaries by failing to distinguish between the type of effect of policies. For example, Anne Schneider and Helen Ingram, "Social Construction of Target Populations: Implications for Politics and Policy," *American Political Science Review* 8, 2 (1993): 334–347.

[17]Edward A. Suchman, *Evaluative Research: Principles and Practice in Public Service and Social Action Programs* (New York: Russell Sage Foundation, 1969), p. 61.

TABLE 6.1

Inputs, Activities, Outputs, Outcomes, and Impacts in Three Issue Areas

Issue Area	*Inputs*	*Activities*	*Outputs*	*Outcomes/Impacts*
CRIMINAL JUSTICE	Expenditures for salaries, computers, maintenance	Arrests Police training Patrols	Criminals arrested per 100,000 known crimes	Criminals convicted per 100,000 known crimes. Safe communities
MUNICIPAL SERVICES	Expenditures for sanitation workers and equipment	Training and promotion of workers	Total residences served	Tons of refuse collected. Clean streets
SOCIAL WELFARE	Number of social workers	Meetings with welfare recipients	Welfare cases per social worker	Standard of living of adults living with dependent children. Self-sufficiency

An *operational definition*, which follows from the constitutive definition, describes the operations required to observe and measure the constitutively defined construct. An *operational definition* of "budget" is a proposed or formally appropriated set of revenues and expenditures, *as observed and measured with a coding instrument that records the magnitude of revenues and expenditures appropriated each fiscal year*. A *variable*, derived from operationally defined constructs, is any characteristic of a person, event, action, or object that takes on different numerical values within a range (a *constant* is a characteristic that does not vary). Examples of policy outcome and impact variables include educational opportunity, public safety, and air quality. Because we frequently lack reliable and valid definitions of these and other variables, a number of appropriate techniques have been developed for assessing the reliability and validity of scales, indexes, tests, questionnaires, and interview schedules.[18]

There are alternative operational definitions of the same variable. This creates problems of interpretation. For example, it may be difficult to know whether one, several, or all of the following indicators are adequate measures of the impact of crime-control policies: arrests per officer, arrests per known crime, the ratio of false to total arrests, number of crimes cleared, number of convictions, numbers of citizens reporting victimization by criminals, citizens' feelings of personal safety. Because the relationship between variables and indicators is complex, it is often desirable to use multiple indicators of the same outcome variable.[19] Sometimes it is possible to construct an *index,* that is, a combination of several indicators that together provide a better measure of outcomes and impacts than

[18]Richard A. Zeller and Edward G. Carmines, *Measurement in the Social Sciences: The Link between Theory and Data* (New York: Cambridge University Press Archive, 1980); Kerlinger and Lee, *Foundations of Behavioral Research*, 4th ed., pp. 621–688.

[19]The general case for multiple indicators may be found in Eugene Webb and others, *Unobtrusive Measures: Nonreactive Research in the Behavioral Sciences* (Chicago, IL: Rand McNally, 1966). For applications to criminal justice and other areas, see Roger B. Parks, "Complementary Measures of Police Performance," in *Public Policy Evaluation*, ed. Kenneth M. Dolbeare (Beverly Hills, CA: Sage Publications, 1975), pp. 185–218.

any one or two indicators do themselves. Among the many types of indexes used in policy analysis are indices of cost of living, pollution, crime severity, energy consumption, health care status, administrative centralization, social and economic well-being, and the quality of working life.[20]

APPROACHES TO MONITORING

Monitoring may be broken down into several identifiable approaches: social systems accounting, social auditing, policy experimentation, research and practice synthesis, meta-analysis, and case studies. These approaches may be contrasted in terms of two major properties (Table 6.2):

1. *Types of controls.* Approaches to monitoring differ in terms of the control they exercise over policy and program inputs, activities, and outputs. Only one of the approaches (policy experimentation) involves prospective direct controls. The other three approaches exercise control statistically, by measuring retrospectively how much of the variation in outcomes and impacts can be explained by variations in inputs, activities, outputs, as well as factors that are external to the experiment. There is not one type of social experimentation, but several: the randomized controlled trial or randomized experiment, the quasi-experiment, the regression-discontinuity experiment, and the natural experiment.[21]
2. *Type of information required.* Approaches to monitoring differ according to their respective information requirements. Some approaches (policy experimentation and social auditing) require the collection of new information. Social systems accounting may or may not require new information, while research and practice synthesis and meta-analysis rely exclusively on available information. Case studies rely on newly collected and available information.

---

[20]On various kinds of indices, see references in Miller and Salkind, *Handbook of Research Design and Social Measurement,* Part IV; James L. Price, *Handbook of Organizational Measurement* (Lexington, MA: D. C. Heath and Company, 1972); Dale G. Lake and others, *Measuring Human Behavior* (New York: Columbia University Teachers College Press, 1973); Paul N. Cheremisinoff, ed., *Industrial Pollution Control: Measurement and Instrumentation* (Westport, CT: Techonomic Press, 1976); Leo G. Reader and others, *Handbook of Scales and Indices of Health Behavior* (Pacific Palisades, CA: Goodyear Publishing, 1976); Lou E. Davis and Albert B. Cherns, *The Quality of Working Life*, vol. I (New York: Free Press, 1975); and Albert D. Biderman and Thomas F. Drury, eds., *Measuring Work Quality for Social Reporting* (New York: Wiley, 1976). Data appropriate for building scales and indices are at www.data.gov.

[21]Many randomized controlled trials are performed and reported by organizations including the World Bank and the Institute for Educational Science of the U.S. Department of Education. Other resources include the Consortium for Evidence Based Policy, the Campbell Collaboration, and the Cochrane Collaboration. All these sources have web pages and URLs. Three books provide excellent multidisciplinary coverage of all types of social experiments: Shadish, Cook, and Campbell, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*; P. J. Gertler, S. Martinez, P. Premand, L. B. Rawlings, and C. M. Vermeersch. *Impact Evaluation in Practice* (World Bank Publications, 2011); and Joshua Angrist and Jorn-Steffen Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion* (Princeton: Princeton University Press, 2009).

TABLE 6.2

Contrasts among Six Approaches to Monitoring

Approach	Types of Control	Type of Information Required
Social systems accounting	Statistical and graphical	Available information
Social experimentation	Direct manipulation of independent variables Statistical and graphical	New information
Social auditing	Statistical and graphical	New information
Research and practice synthesis	Statistical and graphical	Available information
Meta-analysis	Statistical and graphical	Available information
Case study	Graphical	Available and new information

Each of these six approaches also has other common features. First, each is concerned with monitoring *policy-relevant outcomes and impacts.* In addition, social auditing seeks information about whether the resources/inputs reach the intended beneficiaries. Each approach deals with variables that are relevant to policymakers because they are indicators of policy inputs, activities, outputs, outcomes, and impacts. Some policy-relevant variables can be manipulated by policymakers (e.g., resource inputs or new processes for delivering services), while others cannot. These non-manipulable variables include preconditions, which are present before actions are taken (e.g., the average age or the cultural values of a target group), as well as unforeseen events that occur in the course of policy implementation (e.g., sudden staff turnover, strikes, or natural disasters).

A second common feature of these approaches is that they are *goal-focused.* This means that policy outcomes are monitored because they are believed to enhance the satisfaction of some need, value, or opportunity—that is, outcomes are seen as ways to resolve a policy problem. At the same time, some policy outcomes are monitored because they may inhibit the satisfaction of some need, value, or opportunity. Note also that non-manipulable variables, that is, policy preconditions and unforeseen events, are also goal-focused to the extent that they are known to affect policy outcomes.

A third feature common to these approaches is that they are *change-oriented.* Each approach seeks to monitor change, by analyzing changes in outcomes over time; by comparing such changes across two or more programs, projects, or localities; or by using some combination of the two. While some approaches (social systems accounting) are oriented toward macro level changes in societies, states, regions, and communities, other approaches (social auditing and social experimentation) are primarily oriented toward micro-level changes in programs and projects.

Finally, each approach is concerned with *objective as well as subjective* measures of policy actions and outcomes. For example, all approaches provide measures of such objective outcomes as units of health care received as well as such subjective outcomes as satisfaction with medical and health services. Objective indicators are frequently based on available data (e.g., census materials), while subjective ones are based on new data acquired through sample surveys or interviews. In some instances, objective and subjective measures are based both on available and new information. For example, past studies may be repeated (replicated) to obtain new information that may then be compared with old information in an effort to monitor the direction and pace of social change.[22]

Social Systems Accounting

With this general framework in mind, we can proceed to contrast approaches to monitoring. *Social systems accounting* is an approach and set of methods that permit analysts to monitor changes in objective and subjective social conditions over time.[23] The term *social systems accounting* comes from a report published by the National Commission on Technology, Automation, and Economic Progress, a body established in 1964 to examine the social consequences of technological development and economic growth. The commission's report recommended that the federal government establish a "system of social accounts" that would serve as a counterpart to national economic accounts.[24] Several years earlier, a project undertaken by the American Academy of Arts and Sciences for the National Aeronautics and Space Administration explored the second-order impacts of the space program. One result of this project was the development of methods for monitoring social and political, as well as economic trends, and a major book was published in 1966 under the title *Social Indicators*.[25] While these efforts to develop social systems accounting occurred in the 1960s, they actually represent a continuation of work begun in 1933 under the auspices of the President's Research Committee on Social Trends. Under the direction of sociologist William F. Ogburn, the Committee produced a two-volume report, *Recent Social Trends*, before its work was abandoned during the depression.[26]

Work on social systems accounting continued throughout the 1960s and 1970s. Some of the major products of this period were produced by social scientists interested in objective and subjective indicators of social change.[27] Other major initiatives came from

[22]This combination of old and new information is called *replication of baseline studies*. See Otis Dudley Duncan, *Toward Social Reporting: Next Steps* (New York: Russell Sage Foundation, 1969).

[23]The following account is based on Kenneth C. Land and Seymour Spilerman, eds., *Social Indicator Models* (New York: Russell Sage Foundation, 1974); and Land, "Theories, Models, and Indicators of Social Change," pp. 7–20.

[24]National Commission on Technology, Automation, and Economic Progress, *Technology and the American Economy* (Washington, DC: U.S. Government Printing Office, 1966). A primary influence on the Commission's recommendations was Daniel Bell, whose major work (*Toward Post-Industrial Society: A Venture in Social Forecasting*) was described in Case Study 2.

[25]Raymond A. Bauer, ed., *Social Indicators* (Cambridge, MA: MIT Press, 1966). Chapter 3, by Bertram M. Gross, is titled "The State of the Nation: Social Systems Accounting."

[26]President's Research Committee on Social Trends, *Recent Social Trends* (New York: McGraw-Hill, 1933). Ogburn's major work was *Social Change: With Respect to Culture and Original Nature* (New York: B. W. Huebsch, 1922).

[27]Two companion volumes were devoted, respectively, to objective and subjective dimensions of social change. See Eleanor B. Sheldon and Wilbert E. Moore, eds., *Indicators of Social Change: Concepts and Measurements*

federal agencies, including the Department of Health, Education, and Welfare, the Office of Management and Budget, and the Department of Labor.[28] In addition, international organizations including the United Nations and the Organisation for Economic Co-operation and Development have undertaken work on social indicators, as have the governments of France, the United Kingdom, Germany, Canada, Norway, Sweden, and Japan.[29]

The major analytic element of social systems accounting is the *social indicator*. While there are various definitions of this term, the most useful is one that defines social indicators as "statistics which measure social conditions and changes therein over time for various segments of a population. By social conditions, we mean both the external (social and physical) and the internal (subjective and perceptional) contexts of human existence in a given society."[30] Social indicators, as has been suggested, are both objective and subjective, because they help monitor such objective conditions as urbanization, as well as such subjective conditions as political efficacy, alienation, and satisfaction with municipal services. Social indicators are used for monitoring change at the national, as well as state and municipal levels.[31] Social indicators are also available to monitor special aspects of social change such as pollution, health care, and the quality of working life.[32]

The list of representative social indicators presented in Table 6.3 is drawn from a number of major sources and grouped by area. Note that many indicators are designed to monitor objective social states, for example, the number of persons in state mental hospitals. Other indicators depend on subjective responses, for example, persons afraid to walk alone at night. Second, most indicators are expressed in units of time and are cross-classified by various segments of the population, for example, persons afraid to

(New York: Russell Sage Foundation, 1968); and Angus Campbell and Philip E. Converse, eds., *The Human Meaning of Social Change* (New York: Russell Sage Foundation, 1972). For a recent review, see "America in the Seventies: Some Social Indicators," *Annals of the American Academy of Political and Social Science* 435 (January 1978).

[28]Relevant publications are U.S. Department of Health, Education, and Welfare, *Toward a Social Report* (Washington, DC: U.S. Government Printing Office, 1969); Executive Office of the President, Office of Management and Budget, *Social Indicators, 1973* (Washington, DC: U.S. Government Printing Office, 1974); and U.S. Department of Labor, *State Economic and Social Indicators* (Washington, DC: U.S. Government Printing Office, 1973). In 1974 the National Science Foundation funded a Center for the Coordination of Research on Social Indicators, administered under the Social Science Research Council. The Center publishes *Social Indicators Newsletter*.

[29]For a comprehensive annotated bibliography, see Leslie D. Wilcox and others, *Social Indicators and Societal Monitoring: An Annotated Bibliography* (New York: American Elsevier Publishing, 1974). The major international journal is *Social Indicators Research: An International and Interdisciplinary Journal for Quality of Life Measurement*, published since 1974.

[30]Land, "Theories, Models, and Indicators of Social Change," p. 15.

[31]On social indicators at the community level, see Terry N. Clark, "Community Social Indicators: From Analytical Models to Policy Applications," *Urban Affairs Quarterly* 9, 1 (September 1973): 3–36. On urban social indicators, see M. J. Flax, *A Study in Comparative Urban Indicators: Conditions in 18 Large Metropolitan Areas* (Washington, DC: Urban Institute, 1972).

[32]A comprehensive source on policy indicators is Duncan MacRae Jr., *Policy Indicators* (Chapel Hill: University of North Carolina Press, 1985).

TABLE 6.3

Some Representative Social Indicators

Area	Indicator
Health and illness	Persons in state mental hospitals[1]
Public safety	Persons afraid to walk alone at night[2]
Education	High school graduates aged 25 and older[1]
Employment	Labor force participation rate for women[1]
Income	Percent of population below poverty line[1]
Housing	Households living in substandard units[2]
Leisure and recreation	Average annual paid vacation in manufacturing[1]
Population	Actual and projected population[2]
Government and politics	Quality of public administration[3]
Social values and attitudes	Overall life satisfaction and alienation[4]
Social mobility	Change from father's occupation[1]
Physical environment	Air pollution index[5]
Science and technology	Scientific discoveries[6]
Well-Being	Social, health, and general well-being[3]
Governance	Worldwide Governance Indicators (WGI)[7]

Sources:
[1] U.S. Department of Health, Education, and Welfare (1970).
[2] Office of Management and Budget (1974).
[3] Andrews and Withey (2012).
[4] U.S. Department of Labor (1973).
[4] Campbell, Converse, and Rodgers (1976).
[5] Duncan (1969).
[6] National Science Board (Biennial).
[7] The World Bank, Worldwide Governance Indicators, http://info.worldbank.org/governance/wgi/index.aspx#home

walk alone at night are arranged in a time series and cross-classified by race, age, education, and income. Third, some indicators are expressed in terms of present and past experience, while others are stated as goals. For example, in the 1970s, average annual paid vacation in manufacturing was expressed as the 1967 value of 2 weeks and as a 1976–1979 goal of 4 weeks. Finally, many indicators are related to outcomes and impacts of public policies. For example, the labor force participation rate for women aged 35 to 64 may be taken as an indicator of impacts of equal employment opportunity–affirmative action policies, while indices of air pollution may be used to monitor the impact of programs implemented by the Environmental Protection Agency. In short, most social indicators are policy-relevant, goal-focused, change-oriented, cross-classified, and expressive of objective and subjective social conditions.

In using social indicators for purposes of monitoring, it is frequently necessary to make assumptions as to why a social indicator has changed over time. For example, if reported crime is used to monitor the output of crime control policies, official statistics from the Federal Bureau of Investigation (*Uniform Crime Reports*) indicate that reported crimes per 100,000 inhabitants declined by 0.3 percent between 1975 and 1976 but increased by 76.2 percent between 1967 and 1976. To attribute the decrease or increase to criminal justice policies, it is necessary to make the assumption that changes in crime are a consequence of policy actions undertaken by federal, state, and local law enforcement authorities. This assumption is questionable, even when we have accurate data on resource inputs (expenditures for personnel and equipment), once we take into account a more youthful population following the postwar "baby boom." In many cases, however, we are forced to treat the relation between inputs and outputs as a "black box" or *camera obscura,* that is, an unknown dark area symbolizing what we don't know about the causal mechanism converting inputs and activities into outputs and outcomes.

The use of social indicators has several advantages. First, efforts to develop indicators appropriate for monitoring policy outcomes may alert us to areas where there is insufficient information. For example, while there is much information about municipal services, much of this information deals with outputs, that is, the number and types of services provided per capita. Information about the outcomes and impacts of municipal service policies on citizens—for example, impacts measured in terms of satisfaction with transportation, sanitation, and recreational facilities—is inadequate in most cases.[33] Second, when social indicators provide reliable information about the impacts of policies on target groups and beneficiaries, it becomes possible to modify policies and programs. Social indicators also provide information that helps define and structure policy problems and modify existing alternatives. We may find, for example, that the number of high school graduates aged 25 and over has increased, but that the social mobility of the educated population has not changed. In this case, we may wish to restructure the problem of social inequality in a way that places less emphasis on educational opportunity as a vehicle for social mobility.[34]

Social indicators also have various limitations. The choice of certain indicators (e.g., percent of households below the poverty line) reflects particular social values rather than others and may convey the political biases of analysts.[35] Given that policy problems are dynamic, artificial, and subjective, it is doubtful that any social indicator can be totally free of the values of those who develop and apply the indicator for purposes of monitoring. Second, social indicators may not be directly useful to policymakers faced with practical choices. One study of federal policymakers, for example, finds that social indicators are not seen as having great instrumental value.[36] Hence, while social indicators may help to conceptualize, define, and structure problems, they are often so general that they cannot be used to find specific solutions for problems.

---

[33]This is evident in publications such as the Urban Institute, *Measuring the Effectiveness of Basic Municipal Services* (Washington, DC: Urban Institute, 1974).

[34]See Christopher Jencks and others, *Inequality: A Reassessment of the Effect of Family and Schooling in America* (New York: Basic Books, 1972).

[35]Thomas R. Dye, *Understanding Public Policy*, 3rd ed. (Englewood Cliffs, NJ: Prentice Hall, 1978), pp. 324–325.

[36]Nathan Caplan and others, *The Use of Social Science Information by Policy-Makers at the Federal Level* (Ann Arbor: Institute for Social Research, Center for Research on the Utilization of Scientific Knowledge, University of Michigan, 1975).

Most social indicators are based on available data archives about objective social conditions. While it is easier to use available data archives on objective conditions than it is to collect new information about subjective perceptions of these conditions, it may be just as important to monitor subjective conditions as it is to monitor objective ones. For example, reported crimes per 100,000 inhabitants may decrease, yet citizens' sense of personal insecurity may remain the same or even increase, often depending on whether politicians create false and fear-inspiring information, as did the Trump campaign in the 2016 elections in the United States. Similarly, the number of poverty families may decrease without any appreciable change in perceptions of the quality of life.

Finally, social indicators provide little information about the various ways that policy inputs are transformed into policy outcomes. Claims about variations in policy outcomes and impacts are based on correlational data, rather than on knowledge about the causal processes by which inputs are transformed into outcomes and impacts. Policy inputs are measured, policy outcomes are measured, and the two are related by establishing their degree of association. Concurrently, efforts are made to determine whether preconditions and unforeseen events—that is, factors apart from the original policy inputs—enhance or inhibit the production of some outcome. Yet the process of accounting for these non-policy effects occurs after policy outcomes have been produced and is based on the use of statistical controls, that is, techniques that permit analysts to observe the effects of inputs on outputs and outcomes under various statistical conditions. For example, the effects of educational expenditures on school achievement will be analyzed for poor and middle-class families separately.

Policy Experimentation

One consequence of using social indicators is that it may take a very large number of successes and failures to find out what works and why. Such an approach has been associated with random innovation and contrasted with systematic experimentation.[37] *Random innovation* is the process of executing a large number of alternative policies and programs whose inputs are neither standardized nor systematically manipulated. Because there is no direct control over inputs, activities, and outputs, the outcomes and impacts of policies cannot easily be traced back to known sources. By contrast, *policy experimentation* is the process of systematically manipulating policies and programs in a way that permits corrigible answers to questions about the sources of change in policy outcomes. Policy experimentation, of which there are many varieties—for example, randomized controlled trials in laboratory and field settings, quasi-experimentation, regression-discontinuity analysis, panel studies—is advocated as a way to find solutions for social problems by deliberately maximizing the differences of policies and programs and assessing their consequences prior to making large-scale investments in untested programs.[38]

[37]Rivlin, *Systematic Thinking for Social Action*; and Rivlin, "How Can Experiments Be More Useful?" *American Economic Review* 64 (1974): 346–354.

[38]See, for example, Fairweather and Tornatzky, *Experimental Methods for Social Policy Research*.

Policy experimentation is based on procedures used in practical settings involving agriculture, education, and health.[39] Contrary to conventional wisdom, the validity and practical relevance of policy experimentation do not originate in the methods of the classical laboratory experiment, or in the elegant field experiments conducted in the early twentieth century to develop new statistical algorithms in agronomics. This mistaken view is based on a failure to recognize that many social experiments are not mere extensions of the randomized field experiment, a methodology developed for very different purposes by Jerzy Neyman (1894–1981), a mathematical statistician, and Sir Ronald Fisher (1890–1962), a plant geneticist and mathematical statistician. For example, Donald T. Campbell based much of his early work on field experiments in education, including W. A. McCall's *How to Experiment in Education* (1923), which presented views on randomization and research design that anticipated the work of Neyman, Fisher, and Gossett (aka "Student"). Their interests lay primarily in probability distributions and tests of statistical significance, and not as much in improving schools, governments, or economies. The work of McCall was valued because of its practical contributions to knowledge about educational practice.

The normative ideal of social experimentation has been the so-called *true experiment*—namely, an experiment in which an active policy intervention is applied to randomly selected treatment and control groups, with the aim of establishing the causal relevance of the intervention (e.g., a new teaching method). The true experiment is a normative ideal by virtue of its specification of what were believed to be methodologically optimal procedures for testing the validity of causal inferences, inferences that relate one or more treatments, the presumed causes, to one or more treatment outcomes, the presumed effects.

The operational ideal of the policy experiment, however, is the *quasi-experiment*, namely, an experiment conducted outside the laboratory in a field setting where the random assignment of policies to randomly selected persons, groups, or contexts is impractical, illegal, or unethical. Although quasi-experiments may involve random selection of treatment and control groups (but not random assignment of a treatment to these groups), they are designed in and for complex field settings where manifold contingencies lie beyond the control of experimenters.[40] In summary, the social experiment as a form of monitoring has several important characteristics:

- *Direct control over experimental treatments (policy interventions).* Analysts who use social experimentation directly control experimental treatments (policy interventions) and attempt to maximize differences among them in order to

[39]Donald T. Campbell and Julian C. Stanley, *Experimental and Quasi-Experimental Designs for Research* (Ravenio Books, 1963/2015) explicitly reject the kinds of technically complex experimental designs carried out by the agronomist Ronald Fisher, instead opting for field experiments and quasi-experiments in education that were not always based on random assignment. See also Campbell, "Reforms as Experiments," in *Handbook of Evaluation Research*, vol. 1, ed. Elmer I., Struening and Marcia Guttentag (Beverly Hills, CA: Sage Publications, 1965), pp. 71–100. For a contrary view that incorporates qualitative methods and political discourse into Campbell's quasi-experimental framework see William N. Dunn, "Reforms as Arguments." *Science Communication* 3, 3 (1982): 293–326.

[40]Sometimes quasi-experiments are called "natural experiments." However, a natural experiment is one where the effects of an "experiment" are non-deliberate, for example, an earthquake that causes achievement test scores to fall in affected schools.

produce effects that differ as much as possible. One objective is the establishment of a plausible *counterfactual* claim, that is, a claim that the experimental (policy intervention) group would have experienced no change in the absence of the intervention, because no change occurred in the control group.

- *Random assignment.* Potential sources of variation in policy outcomes other than those produced by the experimental treatment are eliminated by randomly selecting members to participate in the policy experiment, by randomly dividing them into experimental and control groups, and by randomly assigning the treatment (policy intervention) to one of these groups. Random assignment minimizes biases in the selection of members of the experimental and control groups, who may respond differently to the experimental treatment. For example, children from middle-class families may respond to a special education program in a more positive way than children from poor families. This would mean that factors other than the program itself are responsible for such outputs as higher reading scores. These selection biases are reduced or eliminated through randomization.

Policy experiments and quasi-experiments have been advocated as ways to monitor the outcomes of public policy since the New Deal years of the 1930s.[41] In the post-World War II period, social experiments have been conducted in many areas of public policy: public health, compensatory education, welfare, criminal justice, drug and alcohol abuse, population control, nutrition, highway safety, and housing.[42] One of the best-known social experiments is the New Jersey–Pennsylvania Graduated Work Incentives Experiment, funded by the Office of Economic Opportunity to answer questions surrounding the reform of the social welfare system in the 1960s. A random sample of able-bodied men aged 15 to 58 from low-income families was selected from three sites in New Jersey (Trenton, Paterson–Passaic, and Jersey City) and one in Pennsylvania (Scranton). In each city, some families received various levels of guaranteed income and tax breaks, while others received none. Altogether, some 1,350 families participated in the experiment.

Critics of welfare reform expected that income supplements and tax breaks (called a *negative income tax*) would induce persons from low-income families to work less. This expectation was not substantiated by the experiment, which indicated that the experimental (income maintenance) and control (no income maintenance) groups did not differ significantly in their employment behavior, as reflected by changes in earnings before and after the experiment. In fact, the earnings of the income maintenance groups showed a slightly higher increase than those of the control groups.

Social experimentation has the potential of showing accurately whether certain policy actions (e.g., income maintenance) result in certain outcomes (e.g., earnings). The capacity of experiments and quasi-experiments to produce valid causal inferences

[41]A. Stephen, "Prospects and Possibilities: The New Deal and New Social Research," *Social Forces* 13 (1935): 515–521. See Robert S. Weiss and Martin Rein, "The Evaluation of Broad-Aim Programs: A Cautionary Case and a Moral," in *Readings in Evaluation Research*, ed. Francis G. Caro (New York: Russell Sage Foundation, 1971), pp. 37–42.

[42]See Carl A. Bennett and Arthur A. Lumsdaine, eds., *Evaluation and Experiment* (New York: Academic Press, 1975), and Fairweather and Tornatzky, *Experimental Methods for Social Policy Research*.

about the effects of policy actions on outcomes and impacts is called *internal validity*. The greater the internal validity, the more confidence we have that policy outcomes and impacts are the consequence of a policy. Threats to validity are special conditions of persons, policies, or contexts, or methodological features of the experiment itself, that compromise or diminish the validity of claims about policy outcomes. In fact, well-designed social experiments are a way to reduce threats to the internal validity of claims about policy outcomes.[43] Among these threats to internal validity, the most important are the following:

1. *History.* A number of unforeseen events can occur between the time a policy is implemented and the point at which outcomes are measured. For example, public disturbances, strikes, or sudden shifts in public opinion may represent plausible rival explanations of observed variations in policy outcomes.
2. *Maturation.* Changes within members of groups, whether these be individuals, families, or larger social units, may exert an independent effect on policy outcomes. For example, with the passage of time, individual attitudes may change, learning may occur, or geographical units may grow or decline. Such maturation may explain policy outcomes.
3. *Instability.* Fluctuations in a time series may produce unstable variations in outcomes that are a consequence of random error or procedures for obtaining information and not a result of policy actions. Because most time series are unstable, causal inferences about the effects of actions on outcomes are often problematic.
4. *Testing.* The very fact of conducting an experiment and measuring outcomes may sensitize members of experimental and control groups to the aims and expectations of the experiment. When this occurs, policy outcomes may be a result of the meaning attributed to policy actions by participants, and not the actions (e.g., income guarantees) themselves.[44]
5. *Instrumentation.* Changes in the system for measuring an outcome, rather than variations in the outcome itself, may be the reason for the success (or failure) of a policy.
6. *Mortality.* Some members of the experimental or control groups may drop out of the experiment before it is completed. This makes it more difficult to make causal inferences. For example, in the New Jersey–Pennsylvania experiment some families broke up before the experiment was over.
7. *Selection.* In many situations random sampling is not possible, thus requiring a quasi-experimental design that does not fully eliminate biases in selecting

[43]For a fuller discussion of threats to internal and other types of validity, see Campbell and Stanley, *Experimental and Quasi-Experimental Designs for Research*, pp. 5–6; and Shadish, Cook, and Campbell, *Experimental and Quasi-Experimental Design for Generalized Causal Inference*.

[44]The effects of interpretation and testing are sometimes known as the "Hawthorne effect," named after a series of experiments conducted in the Western Electric Company's Hawthorne Plant in Chicago, Illinois, in 1927–1931. The output of workers increased after working conditions and other factors (inputs) were changed. Later, output continued to increase even after working conditions (rest pauses, lighting, and so on) were changed back to their original state, largely because the experiment itself provided opportunities for workers' participation. See Paul Blumberg, *Industrial Democracy: The Sociology of Participation* (London: Constable, 1968).

respondents. For example, if one group of children in a quasi-experimental program designed to improve reading skills is drawn primarily from upper-class families where regular reading is more prevalent, the difference between reading scores before and after the program will be less pronounced than with children drawn from lower-class families. This may make the program look less effective than it really is and, in some cases, can even make the experiment look harmful.[45]

8. *Regression artifacts.* When members of experimental and control groups are selected on the basis of extreme characteristics (e.g., poor children selected for an experimental reading program may be high achievers), artificial changes in reading scores may occur as a result of what is known as *regression toward the mean.* Regression toward the mean is the statistical phenomenon where extreme values of some characteristic of a population automatically regress toward or "go back to" the average of the population as a whole. For example, the height of children of very tall or very short parents tends to regress toward the arithmetic mean of all parents. Children of very tall and very short parents tend, respectively, to be shorter and taller than their fathers and mothers.

The internal validity of social experiments and quasi-experiments may be increased through carefully designed research. Generally, such research involves random selection, repeated measures of outcome variables over time, and measurement of the preprogram outcome measures for some of the experimental and control groups, but not others. The latter procedure eliminates the possibility of testing effects for some groups, whose outcome measures after the program can be compared with those of equivalent groups whose outcomes have been measured both before and after receiving inputs.[46]

Policy experimentation is weakest in the area of *external validity*, which refers to the generalizability of causal inferences outside the particular setting in which an experiment is conducted. Important threats to the external validity of claims about policy outputs are similar to those that threaten internal validity, including interpretation, testing, and selection. In addition, other threats are important to the generalizability of claims. The most important of these is artificiality. Conditions under which a social experiment is carried out may be atypical or unrepresentative of conditions elsewhere. For example, conditions in New Jersey and Pennsylvania are different from those present in San Francisco, Seattle, and Anchorage, making generalizations problematic.

Social experimentation is frequently unsuccessful in monitoring policy processes, including patterns of interaction among staff and clients and their changing attitudes and values. Many of the most important policies and programs are so complex that social experimentation results in the oversimplification of policy processes. For example, social experimentation is not well suited to broad-aim programs that involve high levels of

[45]See Donald T. Campbell and A. E. Erlebacher, "How Regression Artifacts in Quasi-experimental Evaluations Can Mistakenly Make Compensatory Education Look Harmful," in *Compensatory Education: A National Debate*, vol. III, ed. J. Hellmuth (New York: Brunner-Mazel, 1970), pp. 185–225.

[46]The design described here is the "Solomon 4-group design." The comprehensive overview of alternative research designs and their role in reducing threats to internal validity is Fred N. Kerlinger, *Foundations of Behavioral Research*, 2nd ed. (New York: Holt, Rinehart and Winston, 1973), pp. 300–377.

conflict among stakeholders, or contexts where the same inputs of personnel, resources, and equipment are perceived in altogether different ways by different groups.[47] For these and other reasons, various qualitative methods for obtaining the subjective judgments of stakeholders have been used to uncover otherwise neglected aspects of policy processes. These methods—which range from participant observation and the use of logs or diaries to group sessions that resemble a policy Delphi—are ways to complement or replace social experimentation.[48]

Social Auditing

One of the limitations of social systems accounting and social experimentation—namely, that both approaches neglect or oversimplify policy processes—is partly overcome in social auditing. *Social auditing* explicitly monitors relations among inputs, activities, outputs, outcomes, and impacts in an effort to trace policy inputs "from the point at which they are disbursed to the point at which they are experienced by the ultimate intended recipient of those resources."[49] Social auditing, which has been used in areas of educational and youth policy by analysts at the RAND Corporation and the National Institute of Education, helps determine whether policy outcomes are a consequence of inadequate policy inputs or a result of processes that divert resources or services from intended target groups and beneficiaries.

The essential differences between social auditing and other approaches to monitoring are most easily seen if we compare social auditing with social systems accounting and social experimentation. Variations of the latter approaches have been used to monitor educational policies in two important studies, *Equality of Educational Opportunity* and the Westinghouse Learning Corporation's evaluation of the Head Start program.[50] In the former case, a variation of social systems accounting, the effectiveness of various levels of school resources was monitored by measuring the relation between resource inputs (teachers, textbooks, school facilities) and outcomes, measured in terms of the achievement levels of students. In the Head Start study, which had several characteristics of a social experiment, the relations between special reading and skill-development activities (inputs) and cognitive and affective skills (outputs) of children exposed to the program were studied in a number of cities. The essential feature of these two approaches is that

[47]See Robert S. Weiss and Martin Rein, "The Evaluation of Broad-Aim Programs: A Cautionary Case and a Moral," in *Readings in Evaluation Research*, ed. Caro, pp. 287–296; and Ward Edwards, Marcia Guttentag, and Kurt Snapper, "A Decision-Theoretic Approach to Evaluation Research," in *Handbook of Evaluation Research*, vol. 1, ed. Struening and Guttentag, pp. 139–182.

[48]See M. G. Trend, "On the Reconciliation of Qualitative and Quantitative Analysis: A Case Study," *Human Organization* 37, 4 (1978): 345–354; and M. Q. Patton, *Alternative Evaluation Research Paradigm* (Grand Forks: University of North Dakota Study Group on Evaluation, 1976).

[49]James S. Coleman, *Policy Research in the Social Sciences* (Morristown, NJ: General Learning Press, 1972), p. 18. See also Michael Q. Patton, *Utilization-Focused Evaluation* (Beverly Hills, CA: Sage Publications, 1978).

[50]See James S. Coleman and others, *Equality of Educational Opportunity* (Washington, DC: U.S. Government Printing Office, 1966): and Victor Cicarelli and others, *The Impact of Head Start: An Evaluation of the Effects of Head Start on Children's Cognitive and Affective Development* (Washington, DC: U.S. Government Printing Office, 1969).

> the policy inputs are measured, policy outcomes are measured, and the two are related (with, of course, the use of experimental or statistical controls to neutralize the effect of situational variables). Whatever institutional structure intervenes is taken in effect as a black box into which the input resources go and out of which the desired outcomes and side effects come. In the first research study we mentioned (*Equality of Educational Opportunity*), the school was the black box. Inputs to it were teacher salaries, pupil–teacher ratio, age of textbooks, size of library, and a variety of other traditional measures of school resources. Outputs studied were achievement of students in verbal and mathematical skills. In the second study, the resource inputs were the additional resources provided by Head Start; the outputs were cognitive and affective measures of the children.[51]

In monitoring policy processes, social auditing provides important information about what goes on inside the "black box." The processes monitored in a social audit are of two main types: resource diversion and transformation.[52] In *resource diversion,* original inputs are diverted from intended target groups and beneficiaries as a result of the passage of resources through the administrative system. For example, the total expenditures for two manpower training programs may be equal, yet one program may expend a higher percentage of funds on salaries and other personnel costs, thus resulting in fewer staff members per dollar and a diversion of services from beneficiaries. Far more important is the process of *resource transformation.* Here resources and their actual receipt by target groups may be identical, yet the *meaning* of these resources to program staff and target groups may be altogether different. If the meaning of resources is in fact different, then resources may have been transformed in such a way as to enhance (or retard) their impact on beneficiaries. For this reason, quantitative methods may be supplemented by *qualitative methods* designed to provide information about the subjective interpretations of policy actions by stakeholders who affect and are affected by the implementation of policies.[53]

A good example of the supplementary use of qualitative methods to monitor policy processes is that provided by a housing experiment conducted in 1972 by the U.S. Department of Housing and Urban Development (HUD). This particular experiment was designed to find out whether direct cash payments for housing would help low-income families to obtain adequate housing on the open market.[54]

In this case, quantitative measures of inputs and outputs not only forced some analysts to treat policy processes as a "black box," initially, it also prevented them from seeing that certain factors—for example, spontaneous assistance by housing counselors—accounted for program impacts. The supplementary use of qualitative methods yielded strikingly different results than those that might have been produced by focusing exclusively on quantitative

[51]Coleman, *Policy Research in the Social Sciences*, p. 18.

[52]The following account of social auditing differs from that of Coleman, which is confined to "resource loss." Coleman's account does not include resource transformation and is based on the restrictive assumption that social auditing is appropriate only for policies involving resource distribution.

[53]Representative expositions of qualitative methodology are Barney G. Glaser and Anselm L. Strauss, *The Discovery of Grounded Theory: Strategies for Qualitative Research* (Chicago, IL: Aldine Publishing, 1967); and Norman K. Denzin, *The Research Act* (Chicago, IL: Aldine Publishing, 1970).

[54]See Trend, "On the Reconciliation of Qualitative and Quantitative Analyses: A Case Study."

measures of inputs and outputs. This case also shows that alternative approaches to monitoring are potentially complementary. Social experimentation and quantitative measurement were successfully combined with social auditing and the qualitative description of policy processes.

Research and Practice Synthesis

Social auditing and social experimentation require the collection of new information. Social systems accounting, while based chiefly on available information, also requires new data insofar as information about subjective social conditions is out of date or unavailable. In contrast to each of these approaches, *research and practice synthesis* is a method of monitoring that involves the systematic compilation, comparison, and assessment of the results of past efforts to implement policies and programs. It has been used to synthesize information in a number of policy issue areas that range from social welfare, agriculture, and education to municipal services and science and technology policy.[55] It has also been employed to assess the quality of policy research conducted on policy outcomes.[56]

There are two primary sources of available information relevant to research and practice syntheses: case studies of policy formulation and implementation, and research reports that address relations among policy interventions and outcomes. When research and practice synthesis is applied to case studies, it may be based on the *case survey method,* a set of procedures used to identify and analyze factors that account for variations in the adoption and implementation of policies.[57] The case survey method requires that the analyst first develop a *case-coding scheme,* a set of categories that capture key aspects of policy inputs, activities, outputs, outcomes, and impacts. In one such case survey, analysts sought to determine the influence of political participation and other process variables on the delivery of municipal services, an outcome variable.[58] Other case surveys, focusing on public agencies and private corporations, have attempted to determine what factors account for successful and unsuccessful efforts to implement and utilize a variety

---

[55]See, for example, Jack Rothman, *Planning and Organizing for Social Change: Action Principles from Social Science Research* (New York: Columbia University Press, 1974); Gerald Zaltman, Robert Duncan, and James Holbek, *Innovations and Organizations* (New York: Wiley-Interscience, 1973); Everett Rogers and F. Floyd Shoemaker, *The Communication of Innovations* (New York: Free Press, 1971); Robert K. Yin and Douglas Yates, *Street-Level Governments: Assessing Decentralization and Urban Services* (Santa Monica, CA: RAND Corporation, October, 1974); U.S. General Accounting Office, *Designing Evaluations* (Washington, DC: U.S. GAO, July 1984); and Richard J. Light and David B. Pillemer, *Summing Up: The Science of Reviewing Research* (Cambridge, MA: Harvard University Press, 1984).

[56]See, for example, Ilene N. Bernstein and Howard E. Freeman, *Academic and Entrepreneurial Research: The Consequences of Diversity in Federal Evaluation Programs* (New York: Russell Sage Foundation, 1975).

[57]See Robert K. Yin and Kenneth Heald, "Using the Case Survey Method to Analyze Policy Studies," *Administrative Science Quarterly* 20, 3 (1975): 371–381; William A. Lucas, *The Case Survey Method: Aggregating Case Experience* (Santa Monica, CA: RAND Corporation, October 1974); and Robert K. Yin, *Case Study Research: Design and Methods*. 5th ed. (Beverly Hills, CA: Sage Publications, 2014).

[58]Yin and Yates, *Street Level Governments*.

TABLE 6.4

Sample Case-Coding Scheme: Representative Indicators and Coding Categories

Type of Indicator	Indicator	Coding Categories
Input	Adequacy of budgetary resources	[] Totally adequate [] Mostly adequate [] Not at all adequate
Activity	Training performance of program staff	[] High test performance [] Medium test performance [] Poor test performance
Output	Number of clients trained by staff	[] 20–25 [] 15–19 [] 10–14 [] None
Outcome	Clients employed within 4 months of job training	[] 9–12 [] 5–8 [] 1–4

of management innovations.[59] Representative indicators and coding categories from a simplified case-coding scheme are illustrated in Table 6.4.

When research and practice synthesis is applied to available research reports, it is based on the *research survey*, research synthesis, or evaluation synthesis, a set of procedures used to compare and appraise results of past research on policy outcomes.[60] Some of the more important applications of the research survey method have yielded insights into factors associated with the diffusion and communication of innovations, the success of planned organizational change efforts, and the achievement of desired outcomes and impacts of policies.[61] The research survey method provides several kinds of information: empirical generalizations about sources of variation in policy outcomes, summary assessments of the confidence researchers have in these generalizations, and policy alternatives or action guidelines that are implied by these generalizations.

[59]William N. Dunn and Frederic W. Swierczek, "Planned Organizational Change: Toward Grounded Theory," *Journal of Applied Behavioral Science* 13, 2 (1977): 135–158. A more recent term for similar procedures is the "realistic research review" developed by R. Pawson, T. Greenhalgh, G. Harvey, and K. Walshe, "Realist Review—A New Method of Systematic Review Designed for Complex Policy Interventions." *Journal of Health Services Research & Policy* 10, Supplement 1 (2005): 21–34.
[60]See Rothman, *Planning and Organizing for Social Change;* U.S. General Accounting Office, *Designing Evaluations*; and Light and Pillemer, *Summing Up*.
[61]Rogers and Shoemaker, *The Communication of Innovations*; Gerald Zaltman, *Toward a Theory of Planned Social Change: A Theory-in-Use Approach,* prepared for the Second Meeting of the Network of Consultants on Knowledge Transfer, Denver, Colorado, December 11–13, 1977; and Rothman, *Planning and Organizing for Social Change,* Ch. 6, pp. 195–278.

The research survey method, like the case survey method, requires the construction of a code book for extracting information from research reports. The typical research report form includes a number of items that help summarize the research and appraise its quality. These items include variables measured, the type of research design and methods used, the issue area in which the research was carried out, and an overall assessment of the reliability and validity of the research findings. A sample research survey form is presented in Table 6.5.

There are several major advantages of research and practice synthesis as an approach to monitoring. First, the case survey and research survey methods are comparatively efficient ways to compile and appraise a large body of cases and research reports on policy

TABLE 6.5

Research Survey Method: Empirical Generalizations, Action Guidelines, and Levels of Confidence in Generalizations

Generalization[1]	Action Guidelines[2]	Confidence in Generalization[3]
1. Current budgetary allocations are marginally different from budgetary allocations last year.	In implementing new policies and programs, analysts should lower expectations where marginal budgetary allocations have been made.	3
2. Positive policy outcomes are associated with the level of wealth of the municipality, state, or region where policies are implemented.	In determining the likely effectiveness of new policies and programs, analysts should focus on wealthy, urban, and industrial areas. Increased tax revenue can then be allocated in poor areas.	4
3. Outputs of industrial policies in municipalities with professional city managers (reform governments) are greater than in municipalities without professional managers (non-reform governments).	Analysts should adjust recommendations according to the presence or absence of a reform government.	2

Notes:
[1] Empirical generalizations are based on available research reports, including books, articles and government reports.
[2] Action guidelines are derived from empirical generalizations. The same generalization frequently implies multiple action guidelines.
[3] The numbers used to express the degree of confidence in a generalization are sometimes based on the number of reliable and valid studies that support the generalization, as determined by the coder.

Source: Adapted from Rothman (1974): 254–265.

implementation.[62] Whether the focus is upon cases or research reports, research and practice synthesis allows analysts to systematically and critically examine different empirical generalizations and their implications. Second, the case survey method is one among several ways to uncover different dimensions of policy processes that affect policy outcomes. Generalizations about policy processes may be used to support conclusions that policies and programs carried out under similar circumstances have similar outcomes. Finally, the case survey method is a good way to examine objective as well as subjective social conditions. The case survey method is a relatively inexpensive and effective way to obtain information about subjective perceptions of policy processes.[63]

The main limitations of research and practice synthesis are related to the reliability and validity of the method. Cases and research reports not only vary in quality and depth of coverage but are often self-confirming. For example, many cases and research reports contain no explicit discussion of the limitations and weaknesses of the study or report, and frequently they put forth only one point of view. Similarly, available cases often report only "successful" efforts to implement policies and programs.[64]

Despite these limitations, research and practice synthesis is a systematic way to accumulate knowledge about policy actions and outcomes. Because it relies exclusively on available information, it is less expensive than social auditing, social experimentation, and social systems accounting. At the same time, these other approaches to monitoring have their own distinctive advantages and limitations. The strengths of one approach are often the weaknesses of another. For this reason, the most persuasive generalizations about policy outcomes are those that have multiple foci (inputs, activities, outputs, outcomes, impacts), use different types of controls (experimental and statistical), and are based on a combination of mixed methods.[65]

Systematic Reviews and Meta-Analyses

The systematic review is an evaluation methodology that summarizes the best available evidence on a specific question using standardized procedures to identify, assess, and synthesize policy-relevant research findings. Systematic reviews investigate the effectiveness of programs and policies through a process that is designed to be accurate, methodologically sound, comprehensive, and unbiased.[66]

The systematic review (SR) and meta-analysis (MA) have become indispensable tools for monitoring the outcomes and impacts of policies. For example, an SR helps identify studies of the outcomes of programs and policies in areas ranging from criminal justice, to education,

[62]On case materials, see the Electronic Hallway, a repository of many hundreds of cases in public administration and policy. This case repository, housed and managed at the Evans School at the University of Washington, is the product of a consortium of public policy schools.

[63]See Zaltman, *Toward a Theory of Planned Social Change*; and Dunn and Swierczek, "Planned Organizational Change: Toward Grounded Theory."

[64]See, for example, U.S. Agency for International Development, Development Experience Clearinghouse (DEC), which archives tens of thousands of case studies and reports (https://dec.usaid.gov/).

[65]See, for example, John W. Creswell and Vicki L. Plano Clark, *Designing and Conducting Mixed Methods Research*. 2nd ed. (Thousand Oaks, CA: Sage Publications, 2011).

[66]Campbell Collaboration Systematic Reviews: Policies and Guidelines (2015): 6. www.campbellcollaboration.org.

policing, and counter-terrorism. In addition, the SR helps evaluate the methodological quality of the studies used to document these outcomes. These two questions—one about the outcomes of programs and the other about the quality of evidence used to make claims about these outcomes—are two sides of the same coin. The coin is evidence of "what works."[67]

The SR and MA (literally the "analysis of analyses") have become centerpieces of evidence-based policy analysis and program evaluation. Tens of thousands of SRs and MAs have been performed in areas including education, juvenile justice, health care, medicine, national security, and international development. The intent of these SRs and MAs is to inform and thereby influence processes of policy design and implementation and to guide future research that may improve these processes. Both also provide essential information for structuring policy problems (Chapter 1). Although the SR and MA are both forms of research synthesis, the main difference between an SR and an MA is that the latter is based entirely on quantitative studies, while the former may involve both quantitative and qualitative studies.[68]

To initiate a new analysis of a significant public policy often requires substantial financial, human, and material resources. Using an example from the National Academy of Sciences, which called for a study of the effect of eating meat on liver cancer, the costs would have approached $1 million. If the study had been done, adding one more study to the 100 that had already been done, a question arose: Would we learn more by conducting the new study, or by investing in an SR or MA to learn as much as possible from the studies already available? Why should we do the 101st policy analysis without knowing what the first 100 concluded?[69]

Methods for SRs and meta-analyses are specialized, and dedicated software is available for storing research studies and analyzing and reporting results.[70] The main steps in conducting an SR or meta-analysis are essentially the same and they are presented in a number of books and manuals.[71] Generally, there are at least seven steps in conducting an SR or MA (Figure 6.2).

1. *Define the preliminary research question*. The research question should be specific enough that it will yield useful studies. For example: Do bundled medical reimbursement systems produce lower costs for the same illness? Avoid overly general questions like "How do medical reimbursement systems affect costs?
2. *Define quality of evidence*. Contemporary information-retrieval systems contain information of all levels of reliability and validity. The quality of evidence that will be accepted into a SR may be "graded" as to quality, for example, by defining the quality

[67]The term "what works" has become a code word for evidence-based program evaluation and policy. The number of new books with "what works" in their titles numbers in the hundreds. The Institute for Education Sciences (IES) of the U.S. Department of Education maintains a What Works Clearinghouse (WWC) in education, including educational policies and programs at https://ies.ed.gov/.

[68]The combination of qualitative and quantitative analysis, sometimes loosely called "mixed-methods," is an opportunity addressed in the work of Ray Pawson in *Evidence-Based Policy: A Realist Perspective* (Thousand Oaks, CA: Sage Publications, 2006).

[69]Richard J. Light and David B. Pillemer discuss this case in *Summing Up: The Science of Reviewing Research* (Cambridge, MA: Harvard University Press, 1984), pp. 6–8.

[70]Software is available from the Cochrane Collaboration (also known as C1) and the Campbell Collaboration (known as C2). One software package is called Review Manager (RevMan) and available from Cochrane. www.cochranecollaboration.org.

[71]For example, Harris M. Cooper, *Research Synthesis and Meta-Analysis: A Step-by-Step Approach*, 5th ed. (Thousand Oaks, CA: Sage Publications, 2017). Light and Pillemer, *Summing Up*, present a similar sequence of steps.

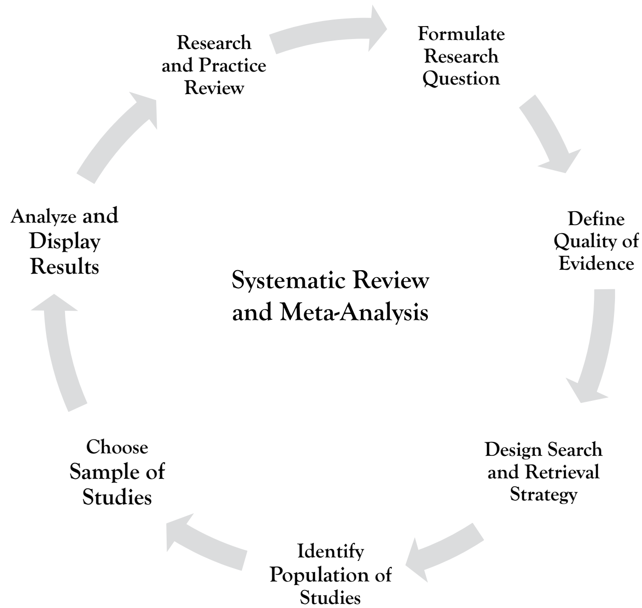


FIGURE 6.2
Steps in Conducting a Systematic Review or Meta-Analysis

of evidence as experimental and quasi-experimental designs and natural experiments, rather than cross-sectional or correlation research or case studies.

3. *Design search and retrieval strategy.* Decide on the key words that will be used for search and retrieval. Depending on the information retrieval system, it is often possible to design a strategy that incorporates in a search string key words at various levels of analysis and of different types. Dates of publication and authors (if known) may be included. Google Scholar and other search engines have a pull-down utility that helps design search and retrieval strategies.
4. *Identify population of studies*. Here all studies that conform to requirements in (2) and (3) will be identified. Often this will mean a population of studies with hundreds of thousands of studies. For example, a search of bundled payment reimbursement and cost-reduction, an innovation included in the Affordable Care and Patient Protection Act ("Obamacare"), yielded 539 citations to studies published after 2000. These studies included research that was published in peer-reviewed journals and studies that were produced by agencies and NGOs but not published. The inclusion of unpublished studies counteracted "publication bias" (the bias that occurs when only peer-reviewed studies with "positive" findings are included).[72]
5. *Choice of relevant sample of studies.* The number of studies that need to be read and coded is related to the budget required to pay coders. If the number of studies numbers in the tens of thousands, which was the case with an important SR of the effectiveness

[72]See pp. 35–37 and 164–165 in Light and Pillemer, *Summing Up*, for a more detailed discussion of publication bias.

of counter-terrorism research,[73] sampling may be necessary. One sampling strategy is a purposive strategy of only including frequently cited articles, which creates publication bias. Another is to randomly sample studies from each page of studies listed.

6. *Analysis and visual display of studies.* The main analyses involve the calculation of effect sizes (e.g., the measured effect of bundled payments on costs), confidence intervals, and funnel plots that may be displayed visually. The analyses involve not only effect sizes but also the effects of the "grade" or quality of the research design.[74]
7. *Researcher and practitioner review.* The purpose of conducting SRs and meta-analyses for purposes of monitoring is to synthesize and evaluate a number of relevant studies on a given policy. Where possible, researchers and practitioners may change the theory initially used to explain the operation of the program, which will change the process of forecasting expected policy outcomes and possibly lead to new policy alternatives. For example, one large SR of studies of the effect of civic education on voting and other forms of political participation concluded that civic education programs, when confined to activities within schools or training facilities, worked poorly. Because some successful programs used actual or simulated participation in external political bodies (e.g., parliamentary committees), new prescriptions based on service learning programs were used as a basis for recommendations.

METHODS FOR MONITORING

Common and specialized methods are available for analyzing and visualizing data obtained with the approaches to monitoring presented in the last section: social systems accounting, social auditing, social experimentation, research and practice synthesis, and systematic reviews and meta-analyses. Methods are matched with these approaches in Table 6.6.

TABLE 6.6

Methods Appropriate to Five Approaches to Monitoring

Approach	Graphic Displays	Tabular Displays	Index Numbers	Effect Sizes	Time-Series Analysis	Regression-Discontinuity Analysis
Social systems accounting	×	×	×	o	×	o
Social auditing	×	×	×	o	×	o
Social experimentation	×	×	×	×	×	×
Research and practice synthesis	×	×	o	×	o	o
Systematic review and meta-analysis	×	×	×	×	o	o

Note: × = method appropriate to approach; o = method not appropriate to approach.

[73]Lum, Cynthia, and Leslie W. Kennedy. *Evidence-Based Counterterrorism Policy* (New York: Springer, 2012).

[74]One of the first efforts to monitor the effects of public programs with a "quality" index was Bernstein, Ilene Nagel, and Howard E. Freeman. *Academic and Entrepreneurial Research: Consequences of Diversity in Federal Evaluation studies* (Russell Sage Foundation, 1975).

Graphic Displays

Information about observed policy outcomes may be presented in the form of *graphic displays*, which are visual representations of the values of one or more output, outcome, and impact variables. A graphic display can be used to depict a single variable at one or more points in time, or to summarize the relation between two variables. In each case, a graph displays a set of data points, each of which may be defined by the coordinates of two numerical scales. The horizontal scale of a graph is called the *abscissa* and the vertical scale is called the *ordinate.* When a graph is used to display a causal relationship, the horizontal axis is reserved for the presumed independent variable (*X*) and called the *X*-axis, while the vertical axis is used to display the presumed dependent variable (*Y*) and called the *Y*-axis. Because one of the main purposes of monitoring is to describe how policies affect policy outcomes, we usually place input, activity, or output variables on the *X*-axis and outcome and impact variables on the *Y*-axis.

One of the most simple and useful kinds of graph is the *time-series graph*, which displays an outcome variable on the vertical axis and time on the horizontal axis. If we wish to monitor the effects of police traffic control activities on motor vehicle deaths between 1970 and 1977, we might use time-series graphs like the ones presented in Figure 6.3(a) and (b). Note that graph (a) has a vertical scale marked off in units of 1,000 motor vehicle deaths and an origin of 44,000 (the origin is the point where the horizontal and vertical axes intersect). By contrast, graph (b) is marked off in units of 5,000 deaths and has an origin of zero. While both graphs display the same data—which show a decrease in deaths after the implementation of the 55 mph speed limit in 1974—the decline in deaths is more dramatic in graph (a) than in graph (b). This example shows how the stretching or compression of

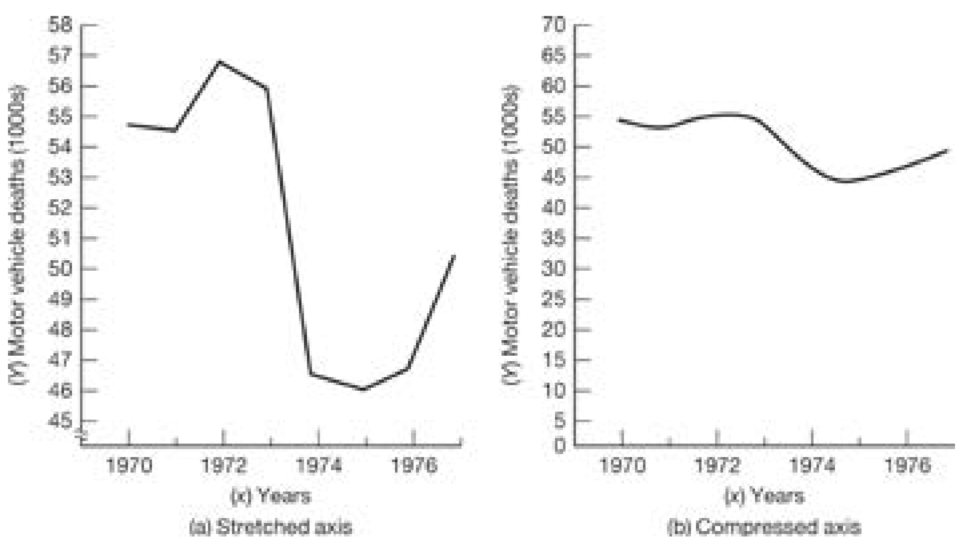


FIGURE 6.3
Two Graphic Displays of Motor Vehicle Deaths

values on the vertical axis can distort the perceived magnitude of data. Because there is no clear rule to follow in selecting the size of intervals, it is essential to examine the underlying meaning of changes depicted in graphs. Another option is to use Edward Tufte's "lie factor," which is based in the relation between spatial and numerical measures in a graph or other visual display.[75]

Graphs may also be used to depict relationships between two or more variables at one point in time. Graphs of this kind, called *scatter diagrams,* were discussed in Chapter 4. We may use a scatter diagram to find out if one variable changes in the same direction as another, that is, whether two variables are correlated. If one of the variables precedes the other in time (e.g., a new program precedes changes in the target population), or if there is a persuasive theory that explains the correlated variables (e.g., economic theory posits that greater income results in a higher propensity to save), then we may be warranted in claiming that there is a *causal relationship* between the variables. Otherwise, variables are simply *correlated*—that is, the values of two variables covary or "go with" each other—and one variable cannot be assumed to be the cause of the other.

A common problem in using graphs is *spurious interpretation*, a situation where two variables that appear to be correlated are actually each correlated with some other variable. A good example of spurious interpretation is that of the analyst who observes from an examination of data on municipal firefighting activities that the number of fire engines present at the scene of a fire (input variable) is positively correlated with the amount of fire damage (output variable). This observed correlation might be used to claim that the number of fire engines deployed does not reduce fire damage, because no matter how many fire engines are present, the amount of fire damage continues to rise at the same rate. This interpretation is spurious (false) because a third variable—the size of fires—is not taken into account. Once we include this variable in our analysis, we find that the size of fires is correlated both with the number of fire trucks deployed and fire damage. The original correlation disappears when we include this third variable in our analysis, and our interpretation is plausible, rather than spurious (Figure 6.4). Spurious interpretation is a serious problem because we can never be completely certain that we have identified all relevant variables that may be affecting the original two-variable relationship.

Another means for displaying data on policy outcomes is the *bar graph,* a pictorial representation that presents data in the form of separated parallel bars arranged along the horizontal (or vertical) axis. The bar graph presented in Figure 6.5 was used to display data on policy inputs (total municipal personnel costs per capita) as part of an analysis of the fiscal crisis of U.S. cities. Note that cities with growing population have the lowest per capita personnel costs, while cities that are losing population have higher costs. The bar graph also shows the somewhat special position of New York City in 1973.

Information about policy outcomes is often available in the form of *grouped frequency distributions,* in which the numbers of persons or target groups in particular categories (e.g., age or income) are presented graphically. For example, if we want to monitor the

[75]Edward R. Tufte, *The Visual Display of Quantitative Information* (Cheshire, CT: Graphics Press, 1983). The lie factor is defined as the size of the effect shown in a graphic divided by the size of the effect in data. A graph showing a ratio of 1/1 = 0.0 indicates no distortion, while a ratio of greater than ± 0.05 signifies substantial distortion. Lie factors of ± 10.0 to ± 50.0 and more are not uncommon.

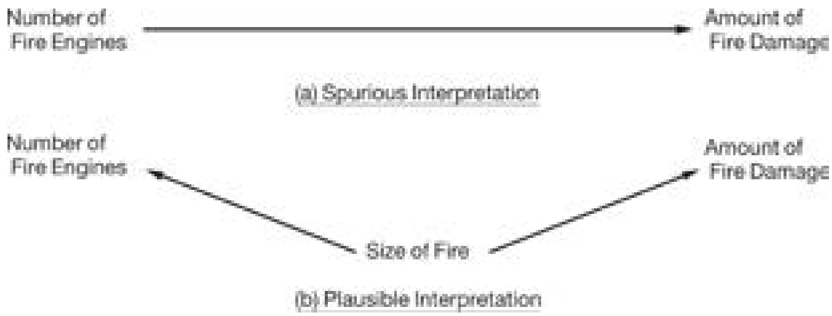


FIGURE 6.4
Spurious and Plausible Interpretations of Data

Source: Adapted from Smith (1975): 325–326.

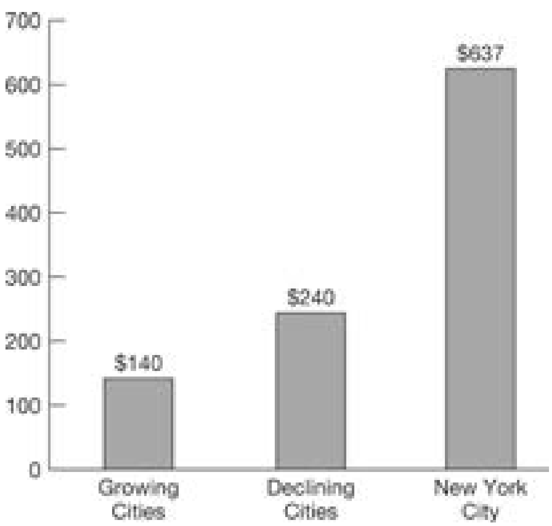


FIGURE 6.5
Bar Graph Showing Municipal Personnel Costs Per Capita for Cities with Growing and Declining Populations and for New York City

Source: Muller (1975).

number of persons below the poverty line in various age groups, we might begin with the data in Table 6.5. These data can then be converted into two kinds of graphic displays: histograms and frequency polygons.

A *histogram* is a kind of bar graph that organizes information about a grouped frequency distribution of an outcome variable at one point in time. In a histogram, the width of the bars along the horizontal axis is equal, and there is no space between the bars. The height of the bars in a histogram shows the frequency of occurrence in each group (called a *class interval*). The histogram illustrated in Figure 6.6(a) shows the numbers of persons below the poverty threshold in several age categories in 1977. The histogram can be converted into a frequency polygon (Figure 6.6[b]) by connecting the midpoints of each class

TABLE 6.7

Grouped Frequency Distribution: Number of Persons below the Poverty Threshold by Age Group in 1977

Age Groups	Number of Persons below the Poverty Threshold
Under 14	7,856,000
14 to 21	4,346,000
22 to 44	5,780,000
45 to 54	1,672,000
55 to 59	944,000
60 to 64	946,000
65 and over	3,177,000
	Total 24,721,000

Note: The poverty threshold is based on an index that takes into account consumption requirements based on sex, age, and residence in farm or nonfarm areas. Poverty thresholds are updated to reflect changes in the Consumer Price Index. In 1977 the poverty threshold for all persons, irrespective of age and place of residence, was $3,067 annually.

Source: U.S. Department of Health and Human Services (2017).

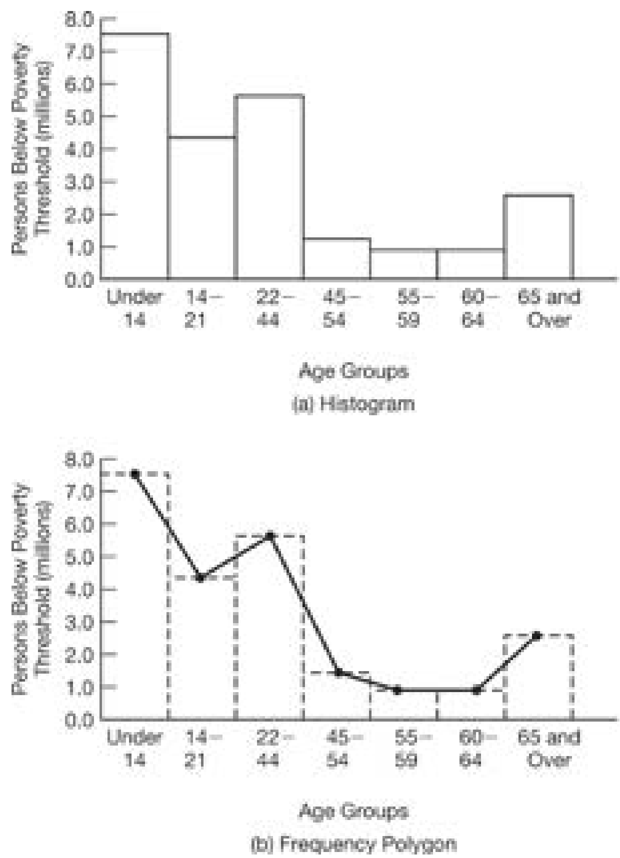


FIGURE 6.6
Histogram and Frequency Polygon: Number of Persons below the Poverty Threshold by Age Group in 1977

Source: U.S. Department of Health and Human Services (2017).

interval with lines. The difference between a histogram and a frequency polygon is that the latter uses lines to represent the frequency distribution.

Another way to display information about policy outcomes is the *cumulative frequency polygon,* a graph that shows the cumulative frequencies of a distribution along the vertical axis. As one moves from left to right along the horizontal axis, the frequency of persons (or families, cities, states) in the first group is plotted on the vertical axis; the frequency of persons in the first and second groups is then plotted; and so on until we reach the end of the horizontal scale, which is the sum of all frequencies. In monitoring the effects of policies designed to alleviate poverty, a useful and widely used type of cumulative frequency polygon is the *Lorenz curve,* which can be used to display the distribution of income, population, or residential segregation in a given population.[76]

A Lorenz curve enables us to compare the share of total income earned by each successive percentage group in the population. These percentage groups are either *quintiles* or *deciles*, which are groups comprised of one-fifth or one-tenth of the population. Figure 6.7 shows the distribution of family personal income in the United States in 1975 and 1989. As the Lorenz curve moves closer to the diagonal (which represents perfect equality), family incomes become more equitably distributed. The curve shows that the distribution of family personal income was more equitable in 1975 than it was in 1989. Since 1989 the distribution of family personal income in the United States has become less and less equitable.

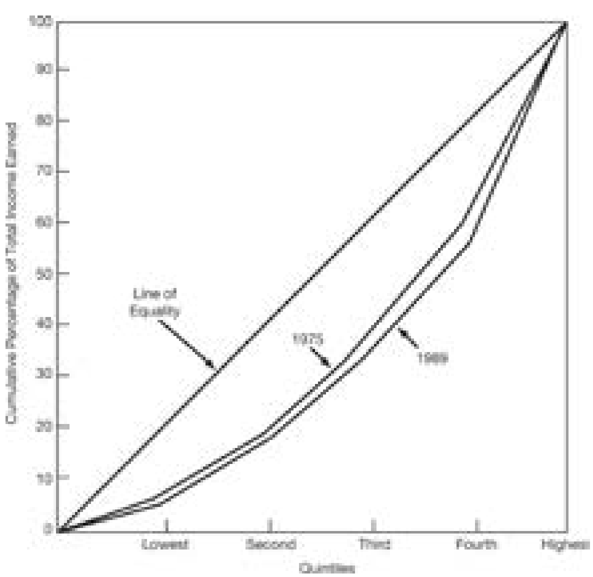


FIGURE 6.7
Lorenz Curve Showing the Distribution of Family Personal Income in the United States in 1975 and 1989

Source: U.S. Department of Health and Human Services (2017).

[76]The original exposition of the Lorenz curve may be found in M. O. Lorenz, "Methods of Measuring the Concentration of Wealth," *Quarterly Publications of the American Statistical Association* 9, 70 (1905): 209–219.

The Gini Index

The Lorenz curve illustrated in Figure 6.7 displays the distribution of income among families at two points in time. Lorenz curves may also be used to display the distribution of certain types of activities (e.g., crime or racial segregation) among spatially organized units such as cities.[77] Lorenz curves can also be expressed in the form of the *Gini concentration ratio* (sometimes called simply the Gini index), which measures the proportion of the total area under the diagonal that lies in the area between the diagonal and the Lorenz curve. The formula for calculating this proportion is

$$\mathrm{GI} = \frac{\left[\left(\sum X_i Y_{i+1}\right) - \left(\sum X_{i+1} Y_i\right)\right]}{\sum X_i Y_{i+1}}$$

where

X_i = the cumulative percentage distribution of the number of areas (e.g., cities)
Y_i = the cumulative percentage distribution of the population or of some activity (e.g., crime)

To illustrate the calculation of the Gini index, let us consider an example from the area of criminal justice policy. Suppose that we wish to depict the concentration of violent crimes in cities of various sizes in 1976. We would use the following procedures, illustrated in Table 6.8.

1. Place the number of cities and the total violent crimes reported for each class-size of cities in columns (1) and (2).
2. Compute the proportionate distribution of cities from column (1) and place in column (3). For example, 59 ÷ 7361 = 0.008.
3. Compute the proportionate distribution of violent crimes from column (2) and place in column (4). For example, 1095 ÷ 2892 = 0.379.
4. Cumulate the proportions of column (3) downward and place the results in column (5). For example, 0 + 0.008 = 0.008, 0.008 + 0.015 = 0.23, and so on.
5. Cumulate the proportions of column (4) downward and place the results in column (6). For example, 0 + 0.379 = 0.379, 0.379 + 0.198 = 0.577, and so on.
6. Multiply the first line in column (6) by the second line in column (5), the second line by the third line, and so on. For example, 0.379 × 0.023 = 0.0087, 0.577 × 0.059 = 0.0340, and so on. Place the results in column (7).
7. Multiply the first line in column (5) by the second line in column (6), and so on. For example, 0.008 × 0.577 = 0.0046, 0.023 × 0.721 = 0.0166, and so on. Place the results in column (8).
8. Sum column (7) and record the total (i.e., 1.3477). Sum column (8) and record the total (i.e., 0.5322).

[77]For other indices of inequality, see Jack P. Gibbs, ed., *Urban Research Methods* (Princeton, NJ: Van Nostrand, 1961).

TABLE 6.8

Computation of Gini Concentration Ratio for Violent Crimes Known to Police per 100,000 Persons in 1976

Size of City	Number of Cities (1)	Violent Crimes (2)	Proportion: Cities (3)	Proportion: Crimes (4)	Cumulative Proportion: Cities (Y_i) (5)	Cumulative Proportion: Crimes (X_i) (6)	$X_i Y_{i+1}$ (7)	$X_{i+1} Y_i$ (8)
250,000 and over	59	1,095	0.008	0.379	0.008	0.379	0.0087	0.0046
100,000–249,999	110	573	0.015	0.198	0.023	0.577	0.0340	0.0166
50,000–99,999	265	416	0.036	0.144	0.059	0.721	0.1016	0.0494
25,000–49,999	604	338	0.082	0.117	0.141	0.838	0.2774	0.1306
10,000–24,999	1,398	254	0.190	0.088	0.331	0.926	0.9260	0.331
Under 10,000	4,925	216	0.669	0.074	1.000	1.000		
Totals	7,361	2,892	1.000	1.000			$\Sigma = 1.3477$	$\Sigma = 0.5322$

$$\text{Gini concentration ratio (GI)} = \frac{[(\Sigma X_i Y_{i+1}) - (\Sigma X_{i+1} Y_i)]}{\Sigma X_i Y_{i+1}}$$

$$= \frac{1.3477 - 0.5322}{1.3477} = \frac{0.8155}{1.3477}$$

$$= 0.605 = 0.61 \text{ (rounded)}$$

Source: Federal Bureau of Investigation (1976).

9. Subtract the total of column (8) from the total of column (7) and then divide by the total of column (7). That is, 1.3477–0.5322 ÷ 1.3477 = 0.8155 ÷ 1.3477 = 0.605 = 0.61. This result is the Gini index.

The Gini index shows the proportion of the area under the diagonal that lies between the diagonal and the Lorenz curve. It can be used to depict the concentration of wealth, income, population, ethnic groups, urban residents, crime, and other social conditions. The advantage of the Gini index is that it is a precise measure of concentration that supplies much more information than the pictorial representation provided by the Lorenz curve. Furthermore, the Gini index ranges from zero (no concentration at all) to 1.0 (maximum concentration). In the example of the concentration of violent crimes, note that approximately 2 percent of the largest cities (100,000 and up) have almost 58 percent of the violent crimes. If violent crimes were more or less equally distributed among cities of all class sizes, the Gini index would be approximately zero. In fact, however, violent crimes are heavily concentrated in large cities. The Gini index value of 0.61 shows that 61 percent of the total area is between the diagonal and the Lorenz curve. This is a substantial concentration with important implications for criminal justice policy in the United States.

Tabular Displays

Another useful way to represent policy outcomes is to construct tabular displays. A *tabular display* is a rectangular array used to summarize the key features of one or more variables. The simplest form of tabular display is the one-dimensional table, which presents information about policy outcomes in terms of a single dimension of interest, for example, age, income, region, or time. In monitoring changes in energy demand in the United States over the period 1950–1970, an analyst might use a one-dimensional tabular display.

Information may also be arranged in tables with two dimensions, for example, changes in poverty rates by age and race over time. An analyst concerned with the impact of welfare reform may wish to monitor changes in the percentage of persons below the poverty level in the United States in 1968, 1990, and 2006. The following table (Table 6.9) provides information showing that the poverty rate among children increased since 1968, while the rate among elders decreased sharply. The overall poverty rate is virtually the same in 2006 as it was in 1968 although the overall rate fell to 11.5 percent by 2016.

Another type of multidimensional table involves the analysis of two or more groups by levels of some outcome variable, which might be employment, services received, or earnings. For example, social experiments are based on comparisons of groups that have received a particular type of treatment (experimental group) and groups that have not (control group). Table 6.10 was used to assess whether groups participating in a publicly funded alcoholism treatment program differed in the number of arrests following treatment (Table 6.10). The results indicate that a court-ordered policy of forced short-term referrals to alcoholism clinics or to Alcoholics Anonymous for treatment (experimental groups) appears to be less effective than no treatment.

Index Numbers

Another useful way to monitor changes in outcome variables over time is index numbers. *Index numbers* are measures of how much the value of an indicator or set of indicators

TABLE 6.9

Poverty Rates in 1968, 1990, and 2006 by Age and Race

	Poverty Rates		
	1968	1990	2006
All	12.8%	13.5%	12.3%
Children	15.4	20.6	17.4
Aged 18–64	9.0	10.7	10.8
Elders	25.0	12.1	9.4
All Less Than Age 65	11.5	13.7	12.7
White	7.5	8.7	8.4
Black	32.8	31.6	24.2
Hispanic	23.8	28.3	20.7

Source: Meyer and Wallace (2009): 9.

TABLE 6.10

Number of Drunk Driving Re-arrests in Treatment and Control Groups

	Treatment Groups			
Re-arrests	*Alcoholism Clinic*	*Alcoholics Anonymous*	*No Treatment*	*Total*
None	26 (32.0)	27 (31.0)	32 (44.0)	85
One or More	56 (68.0)	59 (69.0)	41 (56.0)	156
Total	82 (100.0)	86 (100.0)	73 (100.0)	241

Source: Adapted from Gilbert, Light, and Mosteller (1975): 94.

changes over time relative to a base period. Base periods are arbitrarily defined as having a value of 100, which serves as the standard for comparing subsequent changes in the indicators of interest. Many index numbers are used in public policy analysis. These include index numbers used to monitor changes in consumer prices, industrial production, crime severity, pollution, health care, quality of life, and other important policy outcomes.

Index numbers may focus on changes in prices, quantities, or values. For example, changes in the price of consumer items are summarized in the form of the Consumer Price Index (CPI), while changes in the value of goods produced by industrial firms are summarized with the Index of Industrial Production. Index numbers may be simple or composite. *Simple index numbers* are those that have only one indicator (e.g., the number of crimes per 100,000 persons), while *composite index numbers* include several different indicators. A good example of composite index numbers are two CPIs used to measure the costs of some four hundred goods and services. One index includes all urban consumers, while the other includes urban wage earners and clerical workers.[78]

Index numbers may be *implicitly weighted* or *explicitly weighted*. In the former case, indicators are combined with no explicit procedure for establishing their value. For example, an air pollution index may simply aggregate various kinds of pollutants (carbon monoxide, oxidants, sulfur oxides) without taking into account the different costs in impaired health that result from some of these pollutants (e.g., carbon monoxide). An explicitly weighted index takes such factors into account by establishing the relative value of each indicator.

There are two general procedures for constructing index numbers: aggregation and averaging. Aggregative indices are constructed by summing the values of indicators (e.g., consumer prices) for given periods. Averaging procedures (or the so-called *average of relatives method*) require the calculation of average changes in the value of an indicator over time, while aggregative procedures do not. A useful and simple aggregative price index is the *purchasing power index (PPI)*, which measures the real value of earnings in successive periods. A purchasing power index (PPI) may be used to monitor the impact of wage agreements on the real income of employees as part of public sector collective bargaining

[78]In January 1978, the Consumer Price Index was revised. Since that time, two indices have been produced. A new index includes all urban consumers, irrespective of whether they are employed, and covers 80 percent of the noninstitutional population. The old index includes Urban Wage Earners and Clerical Workers and covers approximately 50 percent of persons in the new index.

TABLE 6.11

Use of Consumer Price Index (1982–1984 = 100) to Compute Purchasing Power Index

	Consumer Price Index (1)	Price Relative (2)	Reciprocal (3)	Purchasing Power Index (4)
2001	177.1	177.1 ÷ 177.1 = 1.0	1.0 ÷ 1.0	× 100 = 100.0
2002	179.9	179.9 ÷ 177.1 = 1.016	1.0 ÷ 1.016	× 100 = 98.4
2003	184.0	184.0 ÷ 177.1 = 1.039	1.0 ÷ 1.039	× 100 = 96.2
2004	188.9	188.9 ÷ 177.1 = 1.039	1.0 ÷ 1.067	× 100 = 93.8
2005	195.3	195.3 ÷ 177.1 = 1.103	1.0 ÷ 1.103	× 100 = 90.6
2006	201.6	201.6 ÷ 177.1 = 1.138	1.0 ÷ 1.138	× 100 = 87.9
2007	207.3	207.3 ÷ 177.1 = 1.171	1.0 ÷ 1.171	× 100 = 85.4
2008	215.3	215.3 ÷ 177.1 = 1.216	1.0 ÷ 1.216	× 100 = 82.3
2009	214.5	214.5 ÷ 177.1 = 1.211	1.0 ÷ 1.211	× 100 = 82.6
2010	218.1	218.1 ÷ 177.1 = 1.232	1.0 ÷ 1.232	× 100 = 81.2

Source: U.S. Department of Labor, Bureau of Labor Statistics.

activities. A purchasing power index is constructed by using values of the CPI. This is done by locating the year (e.g., 2001) for which we wish to determine the purchasing power of wages and converting this value into a *price relative,* that is, a value that expresses the price of goods and services in that year relative to a base year (2001). If we then divide this price relative into one (this is called a *reciprocal*), we will obtain the purchasing power of wages in 2001 dollars.

In Table 6.11, the values of the purchasing power index for given years indicate how much a dollar is worth in real terms, relative to the CPI base year of 2001. For example, in 2005 every dollar in wages buys about 90.6 cents of goods and services when compared with the purchasing power of one dollar in 2001. The purchasing power index can be directly converted into real wages by taking the nominal wages for given years and multiplying by the purchasing power index. This has been done in Table 6.12, which shows the real value of weekly earnings in the years 2001–2010. Real weekly earnings declined from \$596 to \$594 in the period 2001–2008. Real weekly earnings increased to over \$610 in 2009 and over \$606 in 2010.

Another useful aggregative index is the Index of Pollutant Concentration developed under the CAMP of the Environmental Protection Agency. The concentration of various kinds of pollutants, measured in parts per million (ppm) or micrograms per cubic meter of air ($\mu g/m^3$), are punched into a recording tape every five minutes in various locations across the country. Pollutants fall into two main categories: gases and suspended particulates. The maximum concentrations of pollutants reported in Chicago, San Francisco, and Philadelphia for two averaging times (5 minutes and 1 year) in 1964 are displayed in Table 6.13. An averaging time of 1 year is the total quantity of pollutants reported in all 5-minute intervals over a year's period divided by the total number of 5-minute intervals for the year. An averaging time of 5 minutes is simply the maximum concentration of pollutants registered in

TABLE 6.12

Real Weekly Wages in the U.S., 2001–2010

Year	Nominal Earnings (1)	Purchasing Power Index (2)	Reciprocal (3)	Real Wages (4)
2001	$596	1.0	1.000 × $596 =	$596
2002	608	98.4	.984 × 608 =	598.27
2003	620	96.2	.962 × 620 =	596.44
2004	638	93.8	.938 × 638 =	598.44
2005	651	90.6	.906 × 651 =	589.81
2006	671	87.9	.879 × 671 =	589.80
2007	695	85.4	.854 × 695 =	593.53
2008	722	82.3	.823 × 722 =	594.21
2009	739	82.6	.826 × 739 =	610.41
2010	747	81.2	.812 × 747 =	606.56

Source: U.S. Department of Labor, Bureau of Labor Statistics.

any 5-minute period. An implicitly weighted aggregative-quantity index can be developed to monitor the concentration of pollutants over time. The formula for this index is

$$QI_n = \frac{\sum q_n}{q_0} \cdot 100$$

where

QI_n = the quantity index for gaseous pollutants in a given time period n
q_n = the quantity of gaseous pollutants in period n of a time series
q_0 = the quantity of gaseous pollutants in the base period 0 of a time series

On the basis of this formula, we can calculate index numbers to monitor changes in levels of pollutants (measured in millions of tons) between 1970 and 1975. We would begin by selecting 1970 as the base period (q_0) and calculate values of the quantity index for 1973, 1974, and 1975. These values are provided in Table 6.14, which shows that total pollutant emissions have declined since 1970.

The weakness of this implicitly weighted aggregative-quantity index of pollution is that it does not take into account differences in the concentrations of pollutants or their relative damage to health. By contrast, the index developed as part of the Environmental Protection Agency's CAMP provides explicit weights, because it calculates the maximum concentration of pollutants in given periods (averaging times) and relates these concentrations to known health hazards. For example, while the overall quantity of pollutants has declined since 1970, some cities have registered very high levels of pollution in particular periods. Data on these periods are not reflected in the implicitly weighted quantity index illustrated in Table 6.14. Most important, the concentration of different types of pollutants per time

TABLE 6.13

Maximum Concentration of Pollutants Reported in Chicago, Philadelphia, and San Francisco for Averaging Times of 5 Minutes and 1 Year

	Type of Pollutant									
	Gases[1]						*Suspended Particulates*[2]			
City and Averaging Time	Carbon Monoxide	Hydro-Carbons	Nitric Oxide	Nitrogen Dioxide	Nitrogen Oxides	Oxidant	Sulfur Dioxide	Lead	Organic (Benzene Soluble)	Sulfate
Chicago										
5 minutes	64.0	20.0	0.97	0.79	1.12	0.17	1.62	NA	NA	NA
1 year	12.0	3.0	0.10	0.05	0.15	0.02	0.18	0.6	15.1	16.6
Philadelphia										
5 minutes	43.0	17.0	1.35	0.37	1.50	0.25	1.00	NA	NA	NA
1 year	7.0	2.0	0.04	0.04	0.08	0.02	0.09	0.8	12.2	19.8
San Francisco										
5 minutes	47.0	16.0	0.68	0.38	0.91	0.29	0.26	NA	NA	NA
1 year	7.0	3.0	0.09	0.05	0.14	0.02	0.02	NA	9.2	6.2

Notes:
[1] Gaseous pollutants measured in parts per million (ppm) of air in 1964.
[2] Suspended particulates measured in micrograms per cubic meter ($\mu g/m^3$) of air in 1964.

Source: Singpurwalla (1975): 69–71.

TABLE 6.14

Implicitly Weighted Aggregative Quantity Index to Measure Changes in Pollutants in the United States, 1970–1975

	Quantity			
Pollutant	1970	1973	1974	1975
Carbon monoxide	113.7	111.5	104.2	96.2
Sulfur oxides	32.3	32.5	31.7	32.9
Hydrocarbons	33.9	34.0	32.5	30.9
Particulates	26.8	21.9	20.3	18.0
Nitrogen oxides	22.7	25.7	25.0	24.2
	229.4	225.6	213.7	202.2
Quantity index (1970 = 100)	100.0	98.34	93.16	88.14

$$QI_n = \frac{\sum q_n}{\sum q_0} \cdot 100$$

$$QI_{1970} = \frac{\sum 113.7+32.3+33.9+26.8+22.7}{\sum 113.7+32.3+33.9+26.8+22.7} \cdot 100$$

$$= \frac{229.4}{229.4} \cdot 100 = 100.0$$

$$QI_{1973} = \frac{\sum 111.5+32.5+34.0+21.9+25.7}{\sum 113.5+32.3+33.9+26.8+22.7} \cdot 100$$

$$= \frac{225.6}{229.4} \cdot 100 = 98.34$$

Note: In millions of tons of pollutants (gases and particulates).

Source: U.S. Environmental Protection Agency.

period creates different kinds of health hazards. An explicitly weighted quantity index takes into account both the concentration of pollutants per time period (averaging times) and potential health hazards. While we will not construct an explicitly weighted aggregative-quantity index on the basis of averaging times reported by the CAMP, it is important to consider relationships between the duration of exposure to pollutants and damage to health and the environment (Table 6.15).

There are several indexes widely used to monitor changes in policy outcomes. These include the two CPIs already noted previously (one for all urban consumers and the other for urban wage earners and clerical workers); the Producer Price Index (PPI)—before 1978 called the Wholesale Price Index (WPI)—which is used to measure changes in the prices of all commodities produced in the economy; and the Index of Industrial Production (IIP)—published monthly by the Federal Reserve System—which is used to measure changes in the outputs of plants, mines, and utilities companies. The base period for all these indexes is 1967.

TABLE 6.15

Duration of Exposure to Pollutants and Consequent Damage to Health and the Environment

		Pollutant			
Duration of Exposure	*Consequence*	*Carbon Monoxide*	*Oxidants*	*Particulates*	*Sulfur Oxides*
1 second	Sensation				
	Odor	–	+	–	+
	Taste	–	–	–	+
	Eye irritation	–	+	–	–
1 hour	Athletic performance impaired	–	+	–	–
	Visibility reduced	–	+	+	+
8 hours	Judgment impaired	+	–	–	–
	Stress to heart patients	+	–	–	–
1 day	Health impaired	+	–	+	+
4 days	Health impaired	+	–	+	+
1 year	Vegetation damaged	–	–	+	+
	Corrosion	–	–	+	+

Note: A minus (–) indicates no damaging consequences. A plus (+) indicates damaging consequence.

Source: Singpurwalla (1975): 66.

In addition to these indexes, there are those that can be constructed to measure the severity of crime, the quality of life, the quality of health care, and the quality of the environment.

Index numbers have several limitations. First, explicit weighting procedures often lack precision. For example, estimates of the damage caused by different kinds of pollutants are partly based on judgment, and it is difficult to attach dollar costs to resultant health problems. Second, it is difficult to obtain sample data for indexes that are equally meaningful for all groups in society. The CPI, for example, is based on a sample of more than four hundred consumer goods and services purchased by urban residents. The sample is selected monthly from fifty-six urban areas, and data thus obtained are used to construct a composite national index. While the development of the two-part CPI in 1978 makes index numbers more meaningful to retired, elderly, and unemployed persons, as well as to wage earners and clerical workers, the CPI does not reflect price differences in many urban and rural areas. Finally, index numbers do not always reflect qualitative changes in the meaning or significance of index items over time. For example, as we learn more about the effects of various kinds of pollutants (e.g., lead particulates or asbestos), their significance changes. The simple quantity of pollutants may not reflect the changing social value (or disvalue) attached to them in different time periods.

Index numbers may be used to monitor a wide variety of changes. Yet these techniques provide no systematic way of relating changes in policy outcomes to prior policy initiatives. We now consider four methods that permit analysts to make systematic judgments about the effects of policies on policy outcomes: effect size measurement, interrupted time-series analysis, and regression-discontinuity analysis.[79] These procedures are used in conjunction with graphic displays and based on correlation and regression techniques discussed in the chapter on forecasting (Chapter 4). Although two of these procedures (interrupted time-series and control-series analysis) are equally applicable to social systems accounting, social auditing, and social experimentation, three of them (calculation of effect sizes, interrupted time-series analysis, regression-discontinuity analysis) are exclusively applicable to randomized experiments and quasi-experiments.

Effect Sizes

One of the goals of conducting a systematic review (SR) or meta-analysis (MA) is to determine the extent to which programs and policies have achieved desired outcomes and impacts. There is a virtual consensus in the policy analysis and program evaluation communities that the evidence required to determine "what works" is causal in nature. Generally, policy interventions carried out in the form of experiments have several requirements: a presumed cause is actively manipulated to produce a subsequent outcome, there is a correlation between the policy and the outcome, and randomization is used to assign some units (persons, groups, or larger entities) to an intervention group and another to a control group. Random assignment is used to rule out alternative explanations, for example, the explanation that educational level was responsible for graduation rates may be ruled out by randomly selecting the intervention and control groups so that they have approximately the same mean and variance for the variable, years of education. If successfully executed, randomized assignment can justify a counterfactual claim that the intervention (experimental) group would have behaved virtually the same as the control group if the experiment had not been conducted at all. The control group is necessary because the experimental group cannot be "treated" and "not treated" at the same time. Another form of experiment, the quasi-experiment, is virtually identical to the randomized experiment except that random assignment to the intervention group and control group is not possible for logistical, legal, or ethical reasons.

So far, the notion of "cause" has been presented conceptually, without any attempt to show how cause-and-effect can be measured.[80] In this context, there are four frequently

[79]The following discussion is based on the two classic articles by Campbell, "Reforms as Experiments," and Donald T. Campbell and H. Laurence Ross. "The Connecticut Crackdown on Speeding: Time-Series Data in Quasi-Experimental Analysis." *Law and Society Review* (1968): 33–53.

[80]Causality and causal inference are more complex than suggested here. More advanced treatments of causality are Cook and Campbell, *Quasi-Experimentation: Design and Analysis Issues for Field Settings* (Boston: Houghton Mifflin, 1979), Ch.1; Guido W. Imbens and Donald B. Rubin. *Causal Inference in Statistics, Social, and Biomedical Sciences* (Cambridge: Cambridge University Press, 2015); and Pearl, Judea, Madelyn Glymour, and Nicholas P. Jewell. *Causal Inference in Statistics: A Primer* (New York: John Wiley & Sons, 2016). Pearl and others emphasize

used measures of effect size: the odds-ratio (*O-R*), the binary effect size display (*BESD*), the risk-ratio (*R-R*), and the standardized mean difference, also called Cohen's *d*.[81]

Returning to Table 6.10, one of the questions we want to answer is whether the chances of one or more re-arrests are smaller if a person is in the Alcoholism Clinic, compared with no treatment at all. Table 6.16 answers these questions by calculating an odds-ratio (*O-R*). The formula for the odds-ratio is $O\text{-}R = ad \div bc$, where *a, b, c*, and *d* are defined as shown in Table 6.16.

For the Alcoholism Clinic, $O\text{-}R = a \times d \div b \times c = 26 \times 59 \div 27 \times 56 = 1{,}534 \div 1{,}512 = 1.014$. These odds are very nearly even, indicating that it makes virtually no difference to seek treatment in an Alcoholism Clinic as compared with Alcoholics Anonymous. If instead we compare the odds of success with Alcoholics Anonymous versus no treatment (see Table 6.10), we find: $O\text{-}R = a \times d \div b \times c = 27 \times 41 \div 32 \times 59 = 1{,}107 \div 1{,}888 = 0.586$. While, the odds of success with the Alcoholism Clinic versus Alcoholics Anonymous are about even (1.014), the odds of success with Alcoholics Anonymous versus no treatment are only 0.586, considerably less than 1. In interpreting these odds-ratios, however, we should be aware of possible selection bias, if some of those who do not drink alcohol are over-represented in one of the groups. An improved analysis would include randomly selected groups of drinkers before and after the intervention.

Using Table 6.17 as an example, we can calculate another effect size measure, the Binary Effect Size Display (*BESD*): $(a \div a + c) - (b \div b + d) = (26 \div 82) - (27 \div 86) = 0.317$

TABLE 6.16

Calculation of Odds-Ratio (Using Data from Table 6.10)

	Policy	No Policy
Success	a	b
Failure	c	d

a x d / b x c

TABLE 6.17

Calculation of Binary Effect Size Display (*BESD*) and Risk-Ratio (*R-R*)

Outcome/Policy	Alcoholism Clinic	No Treatment
Success/No Re-arrests	26 (0.317)	27 (0.314)
Failure/One or More Re-arrests	56 (0.683)	59 (0.686)
	82 (100.0)	86 (100.0)

that, while the language of statistics is readily available, it must be replaced with another language in order to make causal inferences.

[81]There are other less frequently used measures of effect size including Pearson's *r*, Cramer's *V*, and *Phi*. Generally, *p*-values and other tests of statistical significance and are not used as effect sizes because they are biased by sample size.

– 0.314 = 0.003. The *BESD* of 0.003 indicates that the Alcoholism Clinic has a positive but negligible effect on re-arrests. The risk-ratio (*R-R*) is similar to the *BESD,* except the denominator is not the column totals (a + c or b + d), but the 3rd and 4th cells: *R-R* = (a ÷ c) – (b ÷ d). The *R-R* expresses how risky it is to adopt one policy compared to another. In this case, the risk is virtually the same (0.464 – 0.457 = 0.007).

Perhaps the most widely used effect size measure is Cohen's (d = ES), defined as follows (ES = effect size and G_1 and G_2 are two groups):

$$ES = \frac{\bar{X}_{G1} - \bar{X}_{G2}}{s_{pooled}} \quad s_{pooled} = \sqrt{\frac{s_1^2(n_1-1)+s_1^2(n_2-1)}{n_1+n_2-2}}$$

For purposes of illustration, we will again use the example of the Alcoholism Clinic versus no treatment. However, in this hypothetical example we will assume that there is a continuous measure of success (e.g., counselor rating of readiness for operating a motor vehicle) ranging from 0.0 to 100.0, and that there is a difference of 3.0. In Table 6.18, the second row is marked "average success" and in the parentheses are the variances (s^2) necessary to calculate the pooled standard deviation (see the previous formula).

The value of Cohen's *d* (0.201), because it is standardized, can be used to compare effects in the same unit of measure. It is generally interpreted as:

0.20 = weak effect
0.50 = moderate effect
0.80 = strong effect

In this instance, Cohen's *d* provides a weak effect (0.201) but one that is much larger than the *O-R* (1.014) or the *BESD* (0.003), both of which are negligible.[82] It should be noted that Cohen and others caution against an inflexible interpretation of the coefficients. For

TABLE 6.18

Calculation of Standardized Mean Difference (Cohen's *d*)

Outcome/Policy	Alcoholism Clinic	No Treatment
Average Success	52.1 (221)	49.1 (225)
N	82	86

$$ES = \frac{\bar{X}_{G1} - \bar{X}_{G2}}{s_{pooled}}$$

$$= \frac{52.1 - 49.1}{14.9} = 0.201$$

[82]A full discussion of other effect size measures and their strengths and weaknesses is Mark W. Lipsey and David B. Wilson. *Practical Meta-analysis* (Thousand Oaks, CA: Sage publications, 2001). An Excel-based Practical Effect Size Calculator has been developed by David B. Wilson. It is available at the Campbell Collaboration website:

example, a very low coefficient of $d = 0.11$ may not actually be a "weak" effect in cases where inadequate policy inputs (investment funds) could not be expected to create a moderate or strong effect.

Interrupted Time-Series Analysis

Interrupted time-series analysis is a set of procedures for displaying in graphic and statistical form the effects of policy interventions on policy outcomes. Interrupted time-series analysis is appropriate for problems where an agency initiates some action that is put into effect across an entire jurisdiction or target group, for example, in a particular state or among all families below the poverty threshold. Because policy actions are limited to persons in the jurisdiction or target group, there is no opportunity for comparing policy outcomes across different jurisdictions or among different categories of the target group. In this situation, the only basis of comparison is the record of outcomes for previous years.

The graphing of interrupted time series is a powerful tool for visually assessing the effects of a policy intervention on a measure of some policy outcome (e.g., highway death rates, persons receiving medical assistance, job trainees employed, hospital mortality rates). As Figure 6.8 shows, some policy interventions have genuine effects on policy outcomes, while others do not. Interventions (a) and (b) indicate that the policy intervention is plausibly related to the increase in some valued outcome, because there is a practically significant jump in the value of the measured policy outcome after the intervention. Intervention (b), however, shows that the valued outcome was not durable—it began to "fade" after the fifth time period.

Intervention (c) also lacks durability, but it was also preceded by an extreme high in the measured policy outcome before the intervention. This suggests that there probably was "regression" toward the average value of the time series after the intervention—that is, the increase between the fourth and fifth period probably would have occurred anyway. Intervention (d) shows that the intervention had no effect because there is no difference in the rate of change in the value of the measured outcome variable. Finally, intervention (e) shows a slight increase in the measured policy outcome, but one that was preceded by an increase already in progress. This weak effect is not sufficient to conclude that the intervention was responsible for the increase between the fourth and fifth periods. If data on subsequent time periods were available, the post-intervention increase might prove to lack durability, or might increase at an increasing rate. Here it is simply not possible to know.

Interrupted time-series analysis is one among many approaches to policy experimentation that are called "quasi-experimental." *Quasi-experimental* refers to experiments that have many but not all characteristics of a genuine experiment: random selection of subjects, random assignment of subjects to experimental and control groups, random assignment of a "treatment" (a policy) to experimental and control groups, and measurement of outcomes before and after the experimental treatment (the policy). The only feature that interrupted

www.campbellcollaboration.org. This web-based effect-size calculator is designed for calculating effect sizes for meta-analyses and systematic reviews and for converting one type of effect size (e.g., *O-R*) into another (e.g., *d*).

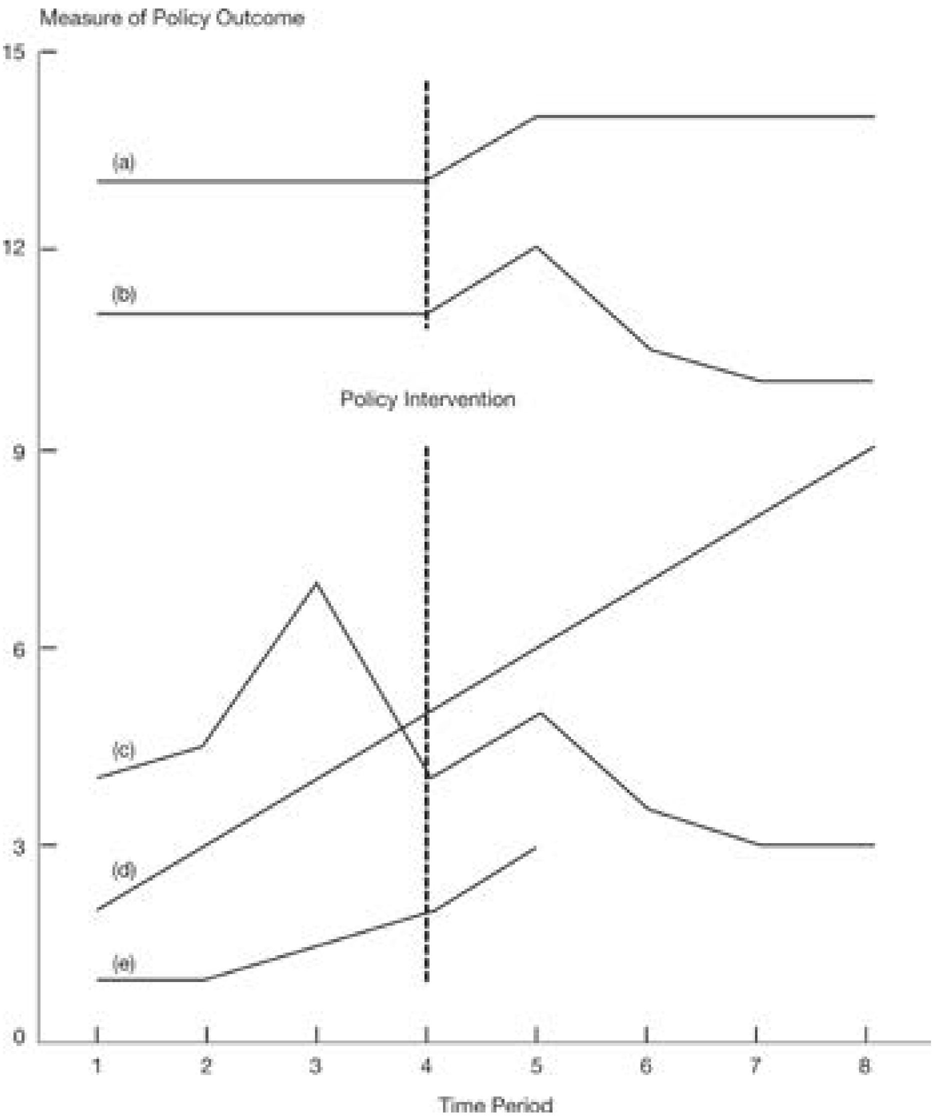


FIGURE 6.8
Interrupted Time Series Showing Effects and No Effects

Source: Campbell and Stanley (1963): 38.

time-series analysis does not have is random selection of participants or subjects. The other three characteristics are present in many quasi-experiments designed to examine policies and programs to determine whether they make a difference in producing some outcome of interest.

The best way to visualize interrupted time-series analysis is to consider a concrete illustration.[83] In 1956, after a record high number of traffic fatalities in 1955, the governor of Connecticut (Abraham Ribicoff) implemented a severe crackdown on violators of the state's speeding laws. At the end of 1956, there were 284 traffic deaths, as compared with 324 deaths the year before. Governor Ribicoff interpreted the outcomes of the crackdown on speeding in the following way: "With the saving of 40 lives in 1956, a reduction of 12.3 percent from the 1955 motor vehicle death toll, we can say that the program is definitely worthwhile." The apparent results of the speeding crackdown are displayed graphically in Figure 6.9. The values on the vertical scale have been deliberately stretched to exaggerate the magnitude of effects.

If we wish to monitor the outcomes of the Connecticut crackdown on speeding using interrupted time-series analysis, we will follow several steps:

1. Compress the values along the vertical axis so that the apparent observed effect is not so dramatic. Here we will use intervals of twenty-five rather than ten deaths.
2. Obtain values of traffic deaths in multiple years before and after the implementation of the policy and plot them on the graph. This creates an extended time series from 1951 through 1959.
3. Draw a vertical line that interrupts the time series at the beginning of the year when the policy was implemented, that is, in 1956 (hence, an "interrupted time series"). Note that intervals used to display years in a time series denote the middle of each

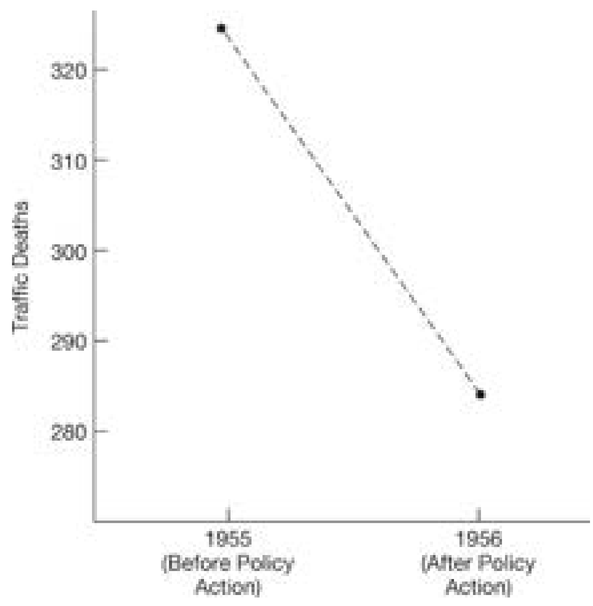


FIGURE 6.9
Connecticut Traffic Deaths before and after the 1956 Crackdown on Speeding

Source: Campbell (1969).

[83]See Campbell, "Reforms as Experiments," pp. 75–81; and Donald T. Campbell and H. Lawrence Ross, "The Connecticut Crackdown on Speeding: Time-Series Data in Quasi-experimental Analysis," *Law and Society Review* 3, 1 (1968): 33–53.

year (July 1), so that any point midway between two intervals is the value of the time series on January 1.

The results of following these procedures are displayed in Figure 6.10, which is called an interrupted time-series graph. Observe that this graph creates an altogether different visual image of policy outcomes than that presented in Figure 6.9.

One of the advantages of interrupted time-series analysis is that it enables a systematic consideration of various threats to the validity of causal inferences about policy outcomes. The importance of these threats to validity (also called rival hypotheses) can be seen when we compare policy outcomes (traffic deaths) displayed in the interrupted time-series graph and Governor Ribicoff's causal inference that the speeding crackdown definitely resulted in lowered traffic fatalities. Let us now examine several threats to the validity of this claim by considering the data displayed in Figure 6.10.

1. *Maturation.* Changes within the members of groups (in this case, drivers) may exert effects on policy outcomes apart from those of policy actions. For example, death rates in Connecticut may have been declining for a number of years as a result of increased social learning among drivers who have been exposed to drivers' education programs or to public safety campaigns. The simple prepolicy–postpolicy graph (Figure 6.9) does not allow us to examine such possible independent effects. The interrupted time-series graph (Figure 6.10) enables us to rule out maturation as a plausible rival explanation, because the time series does not show a long-term secular decline in traffic fatalities before and after the intervention.

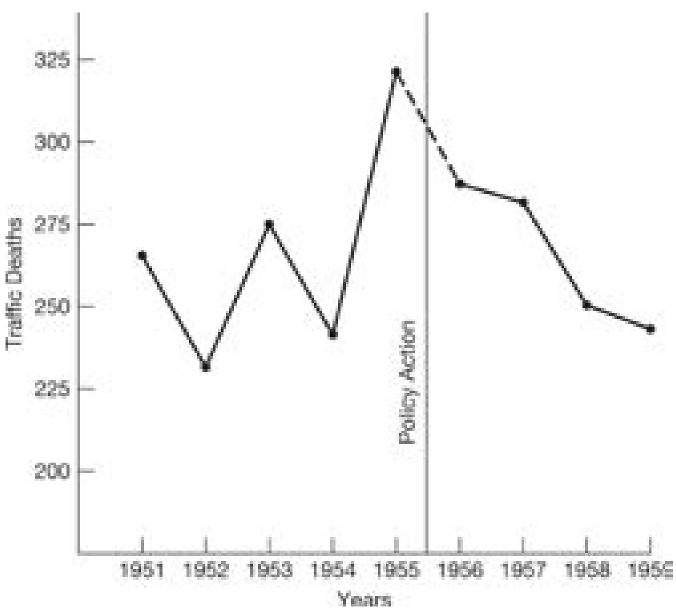


FIGURE 6.10
Extended Time-Series Graph

Source: D. T. Campbell, "Reforms as Experiments," *American Psychologist*, 24, 4 (1969).

2. *Instability.* All time series are to some degree unstable. The interrupted time-series graph permits us to display this instability, which is rather pronounced. The decline between 1955 and 1956 is only slightly greater than declines in 1951–1952 and 1953–1954, suggesting that the policy outcome may be a consequence of instability, and not the policy. On the other hand, there are no increases in traffic fatalities after 1956. Although this lends persuasiveness to the governor's claim, this argument would not have been possible without the interrupted time-series graph of Figure 6.10.
3. *Regression artifacts.* In monitoring policy outcomes, it is often difficult to separate the effects of policy actions from effects exerted by the persons or groups selected for a program. If traffic fatalities for 1955 represent an extreme—as might be the case if we only have Figure 6.10 to go on—then the decline in traffic deaths in 1956 may simply reflect the tendency of extreme scores to regress toward the mean or general trend of a time series. Fatalities in 1955 are in fact the most extreme ones in the series. For this reason, part of the decline from 1955 to 1956 may be attributed to regression artifacts, together with the instability of the time series. Regression artifacts are a particularly important threat to the validity of claims about policy outcomes for one important reason: policymakers tend to take action when problems are most acute or extreme.

Control-series analysis involves the addition of one or more control groups to an interrupted time-series design in order to determine whether characteristics of different groups exert an independent effect on policy outcomes. The logic of control-series analysis is the same as interrupted time-series analysis. The only difference is that a group or groups that were not exposed to the policy actions in question are added to the graph. Figure 6.11 shows the original data from Connecticut along with data on traffic fatalities from control states.

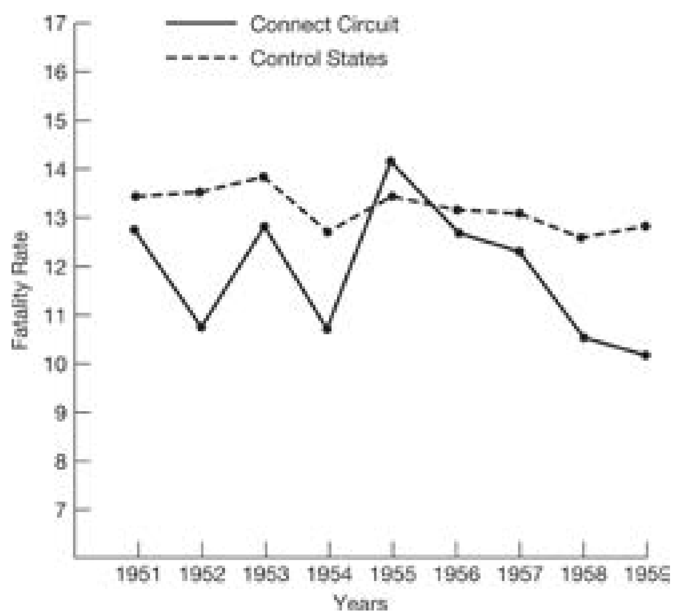


FIGURE 6.11
Control-Series Graph

Source: D. T. Campbell, "Reforms as Experiments," *American Psychologist*, 24, 4 (1969).

Control-series analysis helps further to scrutinize the validity of claims about the effects of policy actions on policy outcomes. For example, rival claims based on history (sudden changes in weather conditions) and maturation (safety habits learned by drivers) would appear to be partly supported by the control-series data. Hence Connecticut and the control states show a similar general pattern of successive increases and decreases in traffic deaths between 1952 and 1955. Yet it is also evident that Connecticut traffic fatalities declined much more rapidly than control states after 1955. The control-series graph not only enables us to assess more critically the governor's claim about the success of his policy, but in this case allows us to rule out plausible rival interpretations. This has the effect of providing much firmer grounds for policy claims.

The use of interrupted time-series and control-series analyses, together with the various threats to internal and external validity, may be viewed as a policy argument. The original data on the decline in traffic fatalities between 1955 and 1956 represent information (*I*), while the policy claim (*C*) is "the program is definitely worthwhile." The qualifier (*Q*) expresses maximum certainty, as indicated by the term "definitely," and the *implicit* warrant (*W*) necessary to carry information to claim is that the policy action caused the reduction in traffic fatalities. What is important in the present case is the series of rebuttals (*R*) that may be advanced on the basis of the various threats to internal and external validity: history, maturation, instability, testing, instrumentation, mortality, selection, regression artifacts. These threats to the validity of claims have been put in the form of a policy argument in Figure 6.12.

The final procedure to be discussed in this chapter is an extension of regression and correlational procedures. Because linear regression and correlation have already been described in some detail, we will not repeat the discussion of these techniques here, except to note that they are applicable to problems of monitoring as well as forecasting. Regression analysis may be used to explain past outcomes, rather than predict future ones, and the results of analysis are sometimes called "post-dictions" to emphasize that we are predicting (in a statistical sense) past events.[84] Even where regression analysis is used to predict future events, the analysis is necessarily based on information acquired by monitoring past policy outcomes.

Imagine that we have obtained information about dollar investments in five manpower training programs, together with information about the subsequent employment of trainees who have completed these programs. The problem can be stated in the following way: How does the level of investment in manpower training programs affect the subsequent employment of trainees? A worksheet necessary to compute the regression equation [$Y_c = a + b(x)$], the coefficient of determination (r^2), the simple correlation coefficient (r), the standard error of estimate ($S_{y.x}$), and an interval estimate (Y_i) is provided in Table 6.19. The SPSS results, displayed in Exhibit 6.1, should be self-explanatory. If they are not, you should return to Chapter 4 and review the section on correlation and regression.

[84] The statistical sense of prediction is not equivalent to a causal interpretation. See J. Pearl, M. Glymour, and N. P. Jewell, *Causal Inference in Statistics: A Primer* (New York: John Wiley & Sons, 2016).

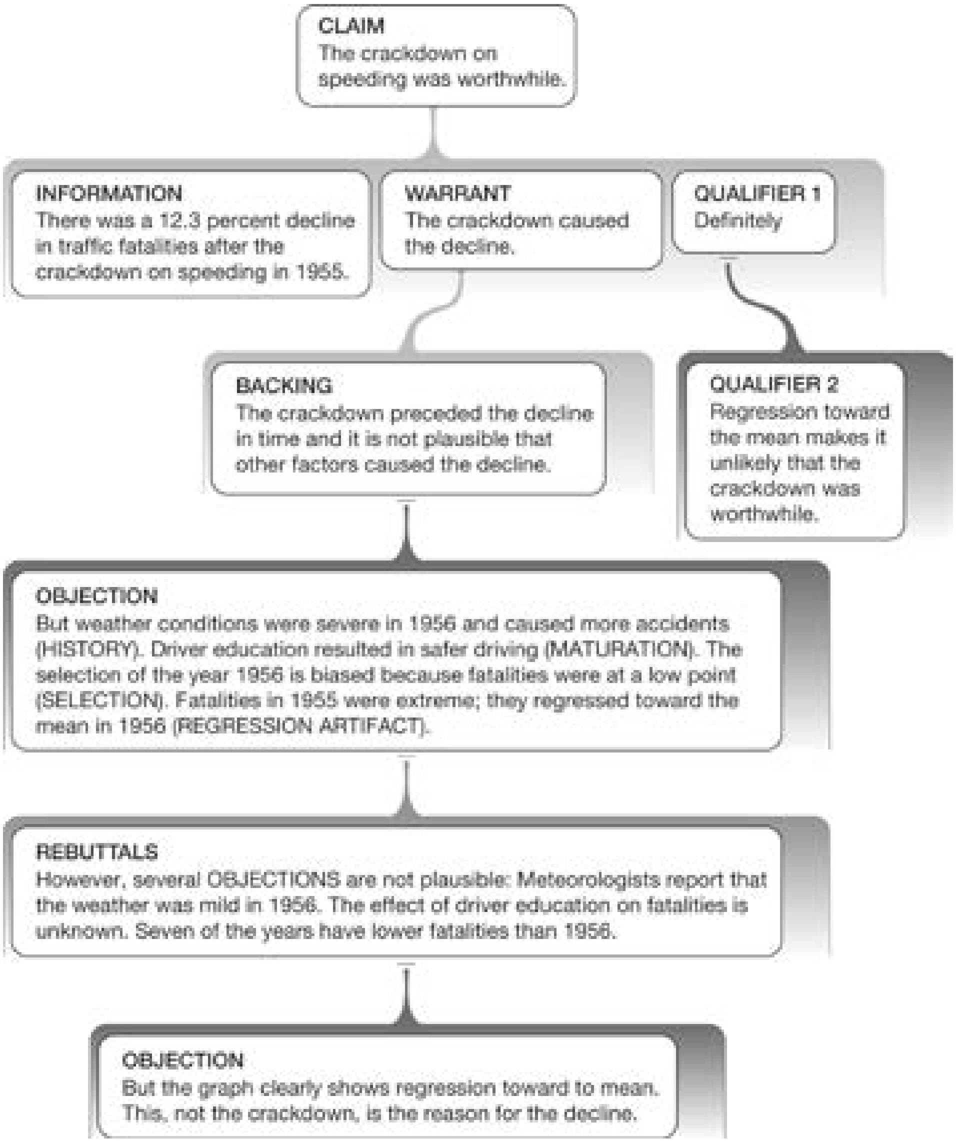


FIGURE 6.12
Threats to Validity as Objections to the Argument That the Speeding Crackdown Was Worthwhile

Regression-Discontinuity Analysis

Having reviewed computational procedures for correlation and regression analysis, we are now in a position to consider regression-discontinuity analysis. *Regression-discontinuity analysis* is a set of graphic and statistical procedures used to compute and compare estimates

TABLE 6.19

Worksheet for Regression and Correlation: Investment in Training Programs and Subsequent Employment of Trainees

Program	Percent of Trainees Employed (Y)	Program Investment (Millions of Dollars)(X)	$Y - \bar{Y}$ (y)	$X - \bar{X}$ (x)	xy	y^2	x^2	Yc	Y – Yc	$(Y - Yc)^2$
A	10	2	–20	–2	40	400	4	12	–2	4
B	30	3	0	–1	0	0	1	21	9	81
C	20	4	–10	0	0	100	0	30	–10	100
D	50	6	20	2	40	400	4	48	2	4
E	40	5	10	1	10	100	1	39	1	1
	150	20	0	0	90	4000	10	150	0	190
	$\bar{Y} = 30$	$\bar{X} = 4$								

$$b = \frac{\Sigma(xy)}{\Sigma(x)^2} = \frac{90}{10} = 9$$

$$S_{yx} = \sqrt{\frac{\Sigma(Y - Yc)^2}{n-2}} = \sqrt{\frac{190}{3}} = 7.96$$

$$a = \bar{Y} - b(\bar{X}) = 30 - 9(4) = -6$$

$$Y_c = a + b(X) = -6 + 9(X) \quad = 48 \pm 15.6 = 63.6 \text{ to } 32.4$$

$$Y_{i(6)}(95\%) = Y_{c(6)} \pm 2Z(s_{y.x}) \qquad = 48 \pm 1.96(7.96)$$

$$r^2 = \frac{b(\Sigma xy)}{\Sigma y^2} = \frac{9(90)}{1000} = 0.81$$

$$r = \sqrt{r^2} = 0.90$$

EXHIBIT 6.1

SPSS Output for Table 6.19

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.900[a]	.810	.747	7.95822426

a. Predictors: (Constant), PROGINV

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	810.000	1	810.000	12.789	.037[a]
	Residual	190.000	3	63.333		
	Total	1000.000	4			

a. Predictors: (Constant), PROGINV
b. Dependent Variable: PERCEMP

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	*t*	Sig.
1	(Constant)	–6.000	10.677		–.562	.613
	PROGINV	9.000	2.517	.900	3.576	.037

a. Dependent Variable: PERCEMP

of the outcomes of policy actions undertaken among two or more groups, one of which is exposed to some policy treatment while the other is not. Regression-discontinuity analysis is the only procedure discussed so far that is appropriate solely for social experimentation. In fact, regression-discontinuity analysis is designed for a particularly important type of social experiment, namely, an experiment where some resource input is in such scarce supply that only a portion of some target population can receive needed resources. Such experiments involve "social ameliorations that are in short supply, and that therefore cannot be given to all individuals."[85] At the same time, randomization may be politically unfeasible and morally unjustifiable, because some of those who are most in need (or most capable) will be eliminated through random selection. For example, limited resources for job training cannot be allocated randomly among a sample of unskilled persons, because job training is intended for the most needy. Similarly, a limited number of scholarships cannot be distributed randomly among a sample of applicants of varying abilities, because it is the most meritorious student (and not the needy one) who is believed to deserve scholarship

[85]Campbell, "Reforms as Experiments," pp. 86–87.

funds. Although randomization would help determine the effects of manpower training or of scholarships on the subsequent behavior of typical or representative members of a target population, randomization frequently violates principles of need or merit and is thus politically and morally unacceptable.

The advantage of regression-discontinuity analysis is that it allows us to monitor the effects of providing a scarce resource to the neediest or most deserving members of a target population that is larger than a given program can accommodate. Imagine a situation where large numbers of deserving persons (e.g., those who obtain high scores on a law school entrance examination) are to be selected to receive some scarce resource (e.g., a scholarship to law school). One of the aims of the admission policy is to select those persons who will succeed later in life, with "success" defined in terms of subsequent income. Only the most deserving applicants are awarded a scholarship, thus satisfying the principle of merit. But the most deserving students would probably be successful in later life even if they received no scholarship. Under these circumstances, it is difficult to determine whether later success in life is a consequence of receiving a scholarship or of other factors, including the family backgrounds of applicants or their social position. Do scholarships affect subsequent success in life?

One way to answer this question is to conduct an experiment (called a "tie-breaking" experiment) where scarce resources are randomly allocated to a small number of persons who are identical or "tied" in their abilities or level of need. The easiest way to visualize a tie-breaking experiment is to imagine five individuals, all of whom have scored 100 on an examination. The problem here is to give awards to two of the most deserving students. Because all students are equally deserving, some procedure must be used to break the tie. One such procedure is randomization.

This same logic is extended to social experimentation through regression-discontinuity analysis. In a tie-breaking experiment, we take a narrow band of merit or need (e.g., as determined by entrance examination scores or family income) above some point that marks the cutoff between those who qualify for some resource and those who do not. To illustrate, imagine that this narrow band of ability is between 90 and 95 percent correct on a law school entrance examination. Persons with scores of 89 percent or less will not be given a scholarship, and persons with scores of 96 percent or more will be given a scholarship. But persons falling into the 90–95 interval (i.e., the narrow band of ability) will be randomly divided into two groups. One group will receive a scholarship and the other will not. Note that this procedure can be justified only under conditions where there are more deserving or needy persons seeking a valued resource than can be accommodated within existing resource constraints.

Without this "tie-breaking equipment," we would be forced to choose only the top students. Under these conditions, we would expect to find entrance examination scores to be strongly and positively correlated with subsequent success in life, as shown by the broken line in Figure 6.13. But by randomly selecting persons from the narrow band of ability (i.e., the 90–95 interval), admitting some and rejecting others (the exact number depends on how many students can be accommodated in any given year), we can find out whether scholarships have an effect on later achievement over and above that of family background and social position. If entrance examination scores do have this augmenting effect, they will be displayed in the form of a discontinuous solid line that separates those who received scholarships from those who did not (Figure 6.13).

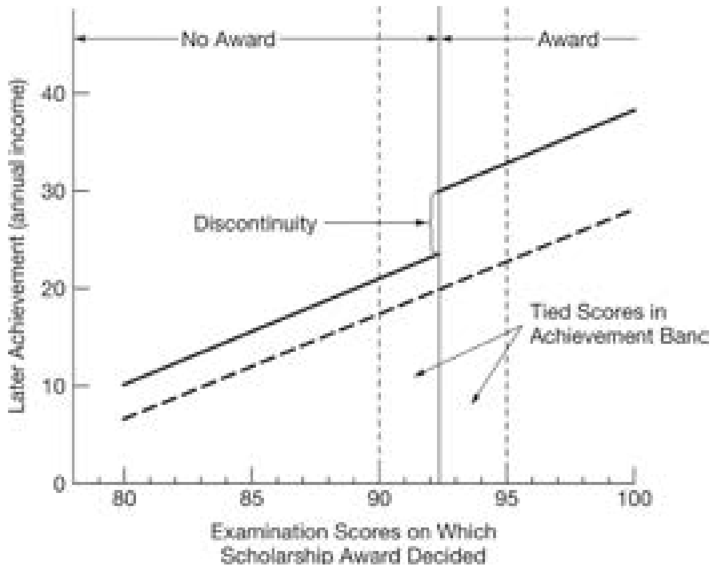


FIGURE 6.13
Tie-Breaking Experiment

Source: D. T. Campbell, "Reforms as Experiments," *American Psychologist,* 24, 4 (1969).

Regression-discontinuity analysis is based on principles of correlation and regression. The main difference is that regression-discontinuity analysis requires that we compute a regression equation [$Y_c = a + b(x)$] for each of two groups, one of which receives some valued resource while the other does not. The standard error of estimate ($S_{y.x}$), the coefficient of determination (r^2), and the simple correlation coefficient (r) may also be computed for each group. Note that the standard errors of regression estimates for the experimental and control groups can be converted into interval estimates that establish the significance of any discontinuity, for example, that displayed by the solid line in Figure 6.13.

Let us now use another hypothetical example to illustrate the application of regression-discontinuity analysis. Imagine that we want to monitor the effects of enforcement policies of the Environmental Protection Agency on the reduction of air pollution levels in various cities. Because enforcement is costly, it is not possible to mount programs in every city where pollution is a significant problem. A random sample of cities is not politically feasible because those areas with the greatest need for enforcement efforts would be eliminated simply by chance. At the same time, we want to find out whether enforcement activities, as compared with no enforcement at all, reduce pollution.

Assume that twenty cities with moderate to high pollution levels have been selected for the enforcement experiment. These cities are ranked according to data on pollutants provided by the CAMP. The distribution of these cities by an index of pollution severity (100 = highest) is illustrated in Table 6.20.

TABLE 6.20

Distribution of Cities by Scores on a Hypothetical Index of Pollution Severity

Pollution Severity		Number of Cities
97		1
95		1
92		1
90		1
89		1
87		1
86		1
85		2
84		1
83	"Tie-breaking" group	1
82		1
81		1
80		1
79		1
78		1
76		1
74		1
71		1
67		1
		Total 20

Source: Fictitious data.

Six cities in the interval between 81 and 85 have been identified as the "tie-breaking" group for purposes of the enforcement experiment. Three of these cities have been selected at random to receive the enforcement program, while the remaining three cities receive no program. Because these cities were chosen solely on the basis of chance, it is unlikely that charges of favoritism will be advanced against policymakers, although pollution severity scores for two of the control cities are slightly above those of the experimental cities. Imagine that air pollution levels in all twenty cities (ten each in the experimental and control groups) were monitored by CAMP one year before and one year after the conclusion of the enforcement experiment. Hypothetical results, displayed in Table 6.21, provide all data necessary to compute regression equations, standard errors of estimates, coefficients of determination, and simple correlation coefficients. Note that these calculations are made separately for the experimental and control groups. The SPSS output is displayed in Exhibit 6.2.

TABLE 6.21

Results of Regression-Discontinuity Analysis for Hypothetical Experimental and Control Cities

			Experimental Cities						
City	*Pre-enforcement Pollution Level (X)*	*Post-enforcement Pollution Level (Y)*	*x*	*y*	x^2	y^2	*xy*	Y_c	$(Y - Y_c)^2$
1	97	93	8.4	8.3	70.56	68.89	69.72	92.56	0.19
2	95	91	6.4	6.3	40.96	39.69	40.32	90.69	0.10
3	92	87	3.4	2.3	11.56	5.29	7.82	87.88	0.77
4	90	87	1.4	2.3	1.96	5.29	3.22	86.01	0.98
5	89	84	0.4	–0.7	0.16	0.49	–0.28	85.07	1.15
6	87	82	–1.6	–2.7	2.56	7.29	4.32	83.20	1.44
7	86	83	–2.6	–1.7	6.76	2.89	4.42	82.27	0.53
8	85	82	–3.6	–2.7	12.96	7.29	9.72	81.33	0.45
9	84	80	–4.6	–4.7	21.16	22.09	21.62	80.39	0.15
10	81	78	–7.6	–6.7	57.76	44.89	50.92	77.59	0.17
	$\bar{X} = 88.6$	$\bar{Y} = 84.7$	0	0	226.40	204.10	211.80	847.00	5.74

$$b = \frac{\Sigma(xy)}{\Sigma(x^2)} = \frac{211.80}{226.40} = 0.936 \qquad a = \bar{Y} - b(\bar{X}) = 84.7 - 0.936(88.6) = 1.8$$

$$Y_C = a + b(X) = 1.8 + 0.936(X)$$

$$S_{y \cdot x} = \sqrt{\frac{(Y - Y_C)^2}{n-2}} = \sqrt{\frac{5.74}{8}} = 0.86$$

$$r^2 = \frac{b(\Sigma xy)}{\Sigma(y^2)} = \frac{936(211.80)}{204.10} = 0.97$$

$$r = \sqrt{r^2} = 0.98$$

Control Cities

City	*Pre-enforcement Pollution Level (X)*	*Post-enforcement Pollution Level (Y)*	x	y	x^2	y^2	xy	Y_c	$(Y-Y_c)^2$
11	85	88	7.5	11.3	56.25	127.69	84.75	87.2	4.84
12	83	84	5.5	7.3	30.25	53.29	40.15	84.4	1.96
13	82	86	4.5	9.3	20.25	86.49	41.85	83.0	1.00
14	80	78	2.5	1.3	6.25	1.69	3.25	80.2	0.04
15	79	75	1.5	−1.7	2.25	2.89	−2.55	78.8	0.04
16	78	77	0.5	0.3	0.25	0.09	0.15	77.4	0.36
17	76	75	−1.5	−1.7	2.25	2.89	2.55	74.6	1.96
18	74	73	−3.5	−3.7	12.25	13.69	12.95	71.8	4.84
19	71	71	−6.5	−5.7	42.25	32.49	37.05	67.6	11.56
20	67	60	−10.5	−16.7	110.25	278.89	175.35	62.0	25.00
	$\bar{X} = 77.5$	$\bar{Y} = 76.7$	0	0	282.50	600.10	395.50	767.0	51.60

$$b = \frac{\Sigma(xy)}{\Sigma(x^2)} = \frac{395.5}{282.5} = 1.4 \qquad a = \bar{Y} - b(\bar{X}) = 76.7 - 1.4(77.5) = -31.8$$

$$Y_C = a + b(X) = -31.8 + 1.4(X)$$

$$S_{y \cdot x} = \sqrt{\frac{(Y - Y_C)^2}{n-2}} = \sqrt{\frac{51.60}{8}} = 2.41$$

$$r^2 = \frac{b(\Sigma xy)}{\Sigma(y^2)} = \frac{1.4(395.5)}{600.10} = 0.923$$

$$r = \sqrt{r^2 = 0.961}$$

EXHIBIT 6.2

SPSS Output for Tables 6.20 and 6.21

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.985[a]	.971	.967	.86302379

a. Predictors: (Constant), PREENF

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	198.142	1	198.142	266.030	.000[a]
	Residual	5.958	8	.745		
	Total	204.100	9			

a. Predictors: (Constant), PREENF
b. Dependent Variable: POSTENF

Coefficients[a]

		Unstandardized Coefficients		Standardized Coefficients		
Model		*B*	Std. Error	Beta	*t*	Sig.
1	(Constant)	1.814	5.089		.356	.731
	PREENF	.936	.057	.985	16.310	.000

a. Dependent Variable: POSTENF

Model Summary

Model	*R*	*R* Square	Adjusted *R* Square	Std. Error of the Estimate
1	.961[a]	.923	.913	2.40831892

a. Predictors: (Constant), PREENFC

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	553.700	1	553.700	95.466	.000[a]
	Residual	46.400	8	5.800		
	Total	600.100	9			

a. Predictors: (Constant), PREENFC
b. Dependent Variable: POSTENFC

EXHIBIT 6.2

SPSS Output for Tables 6.20 and 6.21

Coefficients[a]

		Unstandardized Coefficients	Standardized Coefficients			
Model		*B*	Std. Error	Beta	*t*	Sig.
1	(Constant)	−31.800	11.131		−2.857	.021
	PREENFC	1.400	.143	.961	9.771	.000

a. Dependent Variable: POSTENFC

$$
\begin{aligned}
Y_i\,(\textit{control city } 11) &= Y_{c(85)} \pm 1.96(2.27) \\
&= 87.2 \pm 4.45 \\
&= 82.8 \textit{ to } 91.7
\end{aligned}
$$

The results of the regression-discontinuity analysis help answer several questions about the outcomes of the enforcement experiment. First, observe the pre- and post-enforcement means for the experimental and control cities. The pollution levels for both groups are lower in the post-enforcement period, although the control group received no enforcement program. Nevertheless, the difference between the pre- and post-enforcement means is larger for the experimental cities (88.6 − 84.7 = 3.9) than for the control cities (77.5 − 76.7 = 0.8), suggesting that the enforcement program contributed to the decrease in pollution. Second, experimental cities have a stronger coefficient of determination ($r^2 = 0.971$) and simple correlation coefficient ($r = 0.986$) than those for control cities ($r^2 = 0.932$, $r = 0.961$). The standard error of estimate for the experimental cities ($S_{y \cdot x} = 0.76$) is also lower than that for the control cities ($S_{y \cdot x} = 2.27$). Hence, there is less error in the estimated relationship between enforcement and pollution levels for the experimental cities.

Finally, there is a partial discontinuity between the regression lines for the two groups. If we calculate interval estimates for the two "tie-breaking" cities with identical pre-enforcement scores of 85, we obtain the following results:

$$
\begin{aligned}
Y_i\,(\text{experimental city } 8) &= Y_{c(85)} \pm 1.96(0.76) \\
&= 81.33 \pm 1.49 \\
&= 79.8 \text{ to } 82.8 \\
Y_i\,(\text{control city } 11) &= Y_{c(85)} \pm 1.96(2.27) \\
&= 87.2 \pm 4.45 \\
&= 82.8 \text{ to } 91.7
\end{aligned}
$$

These interval estimates mean that the highest pollution level that is likely to occur 95 percent of the time in experimental city 8 is 82.8, while the lowest that is likely to occur 95 percent of the time in control city 11 is also 82.8. In other words, when we take the highest and lowest pollution levels that are likely to occur 95 percent of the time, the experimental city equals the control city. At the same time, the point estimates are closer for experimental city 10 (77.59) and control city 14 (80.2), indicating that the discontinuity is not as great for some cities.

Note that this kind of comparison would not have been possible without regression-discontinuity analysis. By randomly assigning "tied" scores to experimental and control groups, we are able to determine the effects of the enforcement program on cities that are identical or very similar in their pre-enforcement pollution levels. The results of the regression-discontinuity analysis are more readily visualized if we display them in graphic form. In Figure 6.14, a partial discontinuity between the two groups is evident.

Regression-discontinuity analysis is a useful and powerful technique for monitoring the outcomes of social experiments that involve the distribution of some scarce resource. While our hypothetical example illustrates some of the calculations required to apply the technique, it also avoids some of its complexities. The application of the technique normally requires a much larger number of cases than that used here. For example, as many as 200 cases might have been required to do the kind of analysis just described. The need for a relatively large sample of cases is one of several conditions that must be satisfied in order to obtain valid results. These conditions, which include variables that are normally distributed, are discussed in available statistics texts.[86]

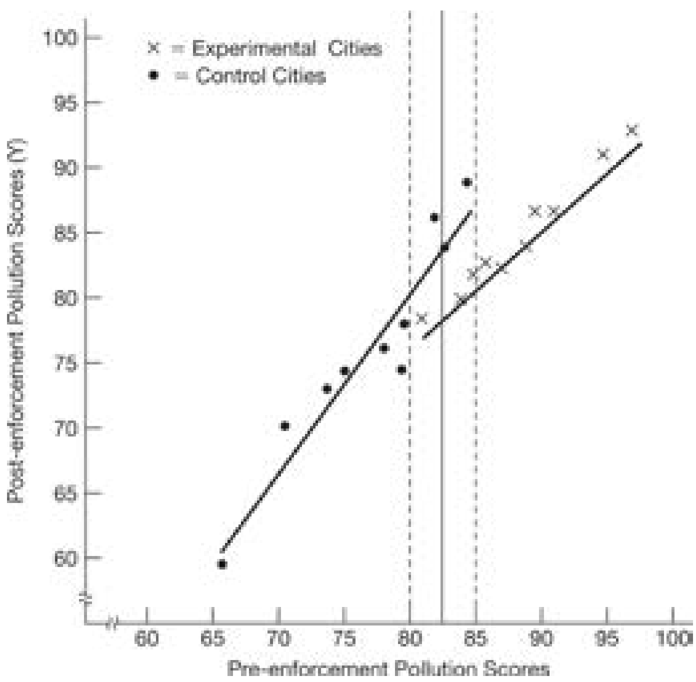


FIGURE 6.14
Results of Regression-Discontinuity Analysis

[86]James H. Stock and Mark W. Watson, *Introduction to Econometrics* (Boston, MA: Addison Wesley, 2003).

CHAPTER SUMMARY

This chapter has provided an overview of the nature and functions of monitoring in policy analysis, described the causal mechanism that links inputs, activities, outputs, outcomes, and impacts, and contrasted six approaches to monitoring. Because a central problem of monitoring is determining whether policies achieve their desired outcomes and impacts, we considered threats to the internal and external validity of causal claims. We concluded with the application of methods for monitoring observed outcomes, which range from social systems accounting and social auditing, to policy experimentation, systematic reviews, and meta-analysis. This should enable you to perform interrupted time-series and control-series analyses to monitor outcomes of national maximum speed limits that were designed to save fuel and lives.

REVIEW QUESTIONS

1. How is monitoring related to forecasting?
2. What is the relationship between ill-structured problems and approaches to monitoring?
3. Construct constitutive and operational definitions for any five of the following variables:

Program expenditure	Equality of educational opportunity
Personnel turnover	National security
Health services	Work incentives
Quality of life	Pollution
Satisfaction with municipal services	Energy consumption
Income distribution	Rapport with clients

4. Listed below are several policy problems. For five of these problems, provide an indicator or index that would help determine whether these problems are being resolved through government action.

Work alienation	School dropouts
Crime	Poverty
Energy crisis	Fiscal crisis
Inflation	Illegal immigration

5. The following table reports the number of criminal offenses known to the police per 100,000 persons. Known offenses are broken down into two categories—total crimes against person and total crimes against property—over the period 1965–1989. Construct two curved-line graphs that display trends in crime rates over the period. Label the two graphs appropriately. What do these graphs suggest about policy outcomes?

TABLE 6.22

Crimes Committed in the United States: Offenses Known to Police per 100,000 Persons, 1985–2009

	1985	1990	1995	2000	2005	2009
Violent	556	730	685	507	469	430
Property	4,651	5,079	4,591	3,618	3,432	3,036

Source: Federal Bureau of Investigation.

6. Mayor Rudy Giuliani often claims credit for the reduction of violent crimes during the period in which he was Mayor of New York City (1994–2001). After examining your graph for violent crimes in the United States, construct an interrupted time-series graph for New York City that shows the start and end of Giuliani's term in office. To what extent is New York City's drop in murders similar to the drop in violent crime in the country as a whole (Question 5)? Why? Does Giuliani deserve credit?

TABLE 6.23

Murders in New York City, 1985–2009

1985	1987	1989	1991	1993	1995	1997	1999	2001	2003	2005	2007
1,384	1,672	1,905	2,154	1,946	1,177	770	671	649	597	539	494

7. In the following table are data on the percentage distribution of family income by quintiles in 1975 and 1989. Use these data to construct two Lorenz curves that depict changes in the distribution of income between 1975 and 1989. Label the two curves and the two axes.

TABLE 6.24

Percentage Distribution of Family Personal Income in the United States by Quintiles, 1975, 1989, 2006, 2015

Quintiles	1975	1989	2006	2015
Highest	41.0	46.7	52.1	51.1
Second	24.0	24.0	20.3	23.2
Third	17.6	15.9	14.3	14.3
Fourth	12.0	9.6	9.5	8.2
Lowest	5.4	3.8	4.7	3.1
Total	100.0	100.0	100.0	100.0

Source: U.S. Bureau of the Census, *Current Population Reports*, Series P-60.

8. Calculate the Gini concentration ratio for 1975 and 2015 data in review question 7. What do the Lorenz curves and Gini coefficients suggest about poverty as a policy problem? If poverty and other problems are "artificial" and "subjective," how valid is the information displayed by the Lorenz curves? Why?
9. Policy issue: Should average monthly benefits paid to mothers under the Aid to Families with Dependent Children (AFDC) program be increased?

Year	Average Monthly Benefit	Consumer Price Index (1982–1984 = 100)
1970	$183.13	38.8
1975	$219.44	53.8
1980	$280.03	82.4
1985	$342.15	107.6
1988	$374.07	118.3

Prepare a policy memo that answers this question.
Before writing the memo:
a. Prepare a purchasing power index for all years in the series, using 1970 as the base year.
b. Convert the average monthly benefits into real benefits for these years.
c. Repeat the same procedures, using 1980 as the base year.

10. Imagine that you are examining the effects of ten scholarship awards on the subsequent performance of disadvantaged youths in college. There are fewer awards than applicants, and awards must be based on merit. Therefore, it is not possible to provide all disadvantaged students with awards, nor is it politically feasible to select students randomly because this conflicts with the principle of merit. You decide to allocate the ten awards according to the following rules: No student will receive an award without scoring at least 86 on the examination; five awards will be given automatically to the top students in the 92–100 interval; and the remaining five awards will be given to a random sample of students in the 86–91 interval.

One year later, you obtain information on the grade point averages of the ten award students and ten other disadvantaged students who did not receive an award. You have college entrance examination scores for all twenty students. Does the provision of scholarship awards to disadvantaged students improve subsequent achievement in college?

Award Group		Nonaward Group	
Examination Scores	*Grade Point Average*	*Examination Scores*	*Grade Point Average*
99	3.5	88	3.2
97	3.9	89	3.2
94	3.4	90	3.3
92	3.0	86	3.0
92	3.3	91	3.5
86	3.1	85	3.1

(*continued*)

Award Group		Nonaward Group	
Examination Scores	*Grade Point Average*	*Examination Scores*	*Grade Point Average*
90	3.3	81	3.0
89	3.2	79	2.8
88	3.2	84	3.0
91	3.4	80	2.6

a. Construct a regression-discontinuity graph that displays examination scores and grade point averages for the award and non-award groups. Use Xs and Os to display data points for the experimental (X) and control (O) groups.
b. Construct a worksheet and compute for each group the values of a and b in the equation $Y_c = a + b(X)$.
c. For each group, write the regression equation that describes the relation between merit (examination scores) and subsequent achievement (grade point averages).
d. Compute the standard error of estimate at the 95 percent estimation interval (i.e., two standard errors) for each group.
e. Compute r^2 and r.
f. Interpret information contained in (a) through (e) and answer the question: Does the provision of scholarship awards to disadvantaged students improve subsequent achievement in college? Justify your answer.

11. Subtract 0.5 from each student's grade point average in the nonaward (control) group.
a. Does the Y intercept change for the control group? Why?
b. Does the slope of the regression line change for the control group? Why?
c. Does the standard error of estimate change for the control group? Why?
d. Do r^2 and r change for the control group? Why?
e. What do your answers show about the appropriateness of correlation and regression analysis for problems where pretest and posttest scores are highly correlated?
f. What, then, are the special advantages of regression-discontinuity analysis?

12. Using at least four threats to validity, construct rebuttals to the following argument: (*B*) The greater the cost of an alternative, the less likely it is that the alternative will be pursued. (*W*) The enforcement of the maximum speed limit of 55 mph increases the costs of exceeding the speed limit. (*I*) The mileage death rate fell from 4.3 to 3.6 deaths per 100 million miles after the implementation of the 55 mph speed limit. (*C*) The 55 mph speed limit (National Maximum Speed Law of 1973) has been definitely successful in saving lives. Study Figure 6.12 before you begin.

DEMONSTRATION EXERCISE

As we know, the simple (bivariate) linear regression equation is written as shown in Equation 6.1.

$$Y = a + b(x)$$

OR

$$y = b_0 + b_1 x \tag{6.1}$$

When we use this equation to estimate the value of a variable in a *time series*, the equation is written as shown in Equation 6.2.

$$y_t = b_0 + b_1 x_t \tag{6.2}$$

Another type of time-series regression equation is one where there are two or more independent (predictor) variables, $X_{1t} \ldots x_{kt}$, which are presumed to be causes of a dependent (response) variable, y_t, which is presumed to be the effect (because regression analysis has nothing to do with causality, per se, the term "presumed" is used here). The equation is written as shown in Equation 6.3.

$$y_t = b_0 + b_1 x_{1t} + b_2 x_{2t} + b_k x_{kt} \tag{6.3}$$

When using time-series regression analysis, it is important to remember the following points, which will be discussed in class:

1. We often want to estimate the effects of a policy intervention on a policy outcome. We do this by creating a so-called "dummy (categorical) variable," which takes the values of 0 before the policy intervention and 1 after the policy intervention. For example, a dummy variable may be symbolized as x_{2t} when the dummy variable is the second predictor variable. The first variable is time, x_{1t}, measured in years. The policy intervention regression equation would be written as shown in Equation 6.4.

$$y_t\,(\text{policy outcome}) = b_0 + b_1 x_{1t}\,(\text{time}) + b_2 x_{2t}\,(\text{policy intervention}) \tag{6.4}$$

2. In linear regression, we must satisfy assumptions of linearity and homoscedasticity. An additional assumption must be satisfied in time-series analysis: The observations of y in the time series must be independent (uncorrelated). This is called the *non-autocorrelation assumption*. Note that, just as there are *tests for linearity* (e.g., plotting the y values against normal scores in a normal probability plot), there are *tests for autocorrelation*. One of these is the Durbin–Watson (D-W) test. We can apply this test with SPSS and other statistical packages.
3. If the autocorrelation coefficient, r, is statistically significant at a specified level of α (usually $p = 0.05$), we reject the null hypothesis that adjacent observations in a time series are uncorrelated—and we thereby accept the alternative hypothesis that adjacent observations *are* autocorrelated. When there is statistically significant autocorrelation, we often can eliminate most of its effects by regressing the values of y_t on their lagged values, y_{t-1}. This is a lag of one time period (e.g., 1 year). The lagged values (one or more time periods) of a variable can be easily computed with SPSS and other statistical packages.

The regression equation with a lagged dependent variable (in this case, 1 year) is written as shown in Equation 6.5.

$$y_t = b_0 + b_1 y_{t-1} \tag{6.5}$$

Equations 6.4 and 6.5 can be combined to express the effects of a lagged policy outcome variable (y_{t-1}), time (x_{t1}), and a policy intervention (x_{t2}) on a policy outcome variable (y_t). Equation 6.6 is the combined equation.

$$y_t = b_0 + b_1 y_{t-1}\,(\text{lag 1 period}) + b_2 x_{2t}\,(\text{time}) + b_3 x_{3t}\,(\text{policy intervention}) \tag{6.6}$$

In this demonstration exercise, we will be using a data file from the Fatal Accident Reporting System of the National Highway Traffic Safety Administration to perform the tasks described here:

1. Estimate the effect of time on fatalities before (1966–1973) and after (1974–2000) the adoption of the 55 mph speed limit. Run two separate regressions. Interpret the two regression estimates. In this part of the exercise, do not use data after 2000.
2. Estimate the effect of employment on fatalities. On the basis of the 1966–1990 series, forecast traffic fatalities for 2009. The actual value of fatalities in 2009 is 33,808. Interpret the computer printout and compare the forecast with actual fatalities for 2009.

3. Estimate the effect of time (years) and the policy intervention (55 mph speed limit) on fatalities, using a dummy variable. Assess whether the effect of the policy intervention is statistically significant and interpret the computer output.
4. Reestimate the equation from the previous task. This time, control for the effect of autocorrelation by adding Y_{t-1} (the lagged value of the dependent variable) as a third predictor. Interpret the output and compare it to output obtained in Step 3.
5. Estimate the effect of the policy intervention on fatalities, after including miles traveled and employment as additional predictor variables. In this regression, do not include time as a predictor variable. Interpret your computer output.
6. Use the data on European traffic fatalities per 100,000 persons to create interrupted time series and control series graphs for (a) the United States (US), (b) for European countries that adopted 48- and 54-mph (80- and 90-kph) speed limits (EUREXP), and (c) for European countries that did not adopt a speed limit (EURCON). Interpret the graphs and indicate how your analysis answers the question: Did the speed limit cause the decline in fatalities?

BIBLIOGRAPHY

Andrews, Frank M., and Stephen B. Withey. *Social Indicators of Well-Being: Americans' Perceptions of Life Quality* (New York: Springer Science & Business Media, 2012).

Campbell, Angus, Philip Converse, and Willard Rodgers. *The Quality of American Life*. New York: Russell Sage Foundation, 1976.

Campbell, Donald T. *Methodology and Epistemology for Social Science: Selected Papers*, ed. E. Samuel Overman. Chicago, IL: University of Chicago Press, 1989.

Campbell, Donald T. "Reforms as Experiments." *American Psychologist*, 24, 4 (1969): 409–429.

Campbell, Donald T. and Julian C. Stanley. *Experimental and Quasi-Experimental Designs for Research*. Chicago, IL: Rand McNally, 1963.

Cook, Thomas D., and Charles S. Reichardt, eds. *Qualitative and Quantitative Methods in Evaluation Research*. Beverly Hills, CA: Sage Publications, 1979.

Cooper, Harris, Larry V. Hedges, and Jeffrey C. Valentine, eds. *The Handbook of Research Synthesis and Meta-Analysis*. New York: Russell Sage Foundation, 2009.

Duncan, Otis Dudley. *Toward Social Reporting: The Next Steps*. New York: Russell Sage Foundation, 1969.

Federal Bureau of Investigation. *Uniform Crime Reports for the United States*. Washington, DC, 1976.

Gilbert, John P., Richard J. Light, and Frederick Mosteller. "Assessing Social Innovations: An Empirical Base for Policy," in *Evaluation and Experiment: Some Critical Issues in Assessing Social Programs,* ed. Carl A. Bennett and Arthur A. Lumsdaine. New York: Academic Press, 1975.

Imbens, Guido W., and Donald B. Rubin. *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge: Cambridge University Press, 2015.

Light, Richard J., and David B. Pillemer. *Summing Up: The Science of Reviewing Research*. Cambridge, MA: Harvard University Press, 1984.

MacRae, Duncan, Jr. *Policy Indicators: Links between Social Science and Public Debate*. Chapel Hill: University of North Carolina Press, 1985.

Meyer, Daniel R., and Geoffrey L. Wallace. "Poverty Levels and Trends in Comparative Perspective." *Focus* 26, 2 (Fall 2009).

Miles, Matthew B., and A. Michael Huberman. *Qualitative Data Analysis: A Sourcebook of New Methods*. Beverly Hills, CA: Sage Publications, 1984.

Muller, Thomas. *Growing and Declining Urban Areas: A Fiscal Comparison*. Washington, DC: Urban Institute, 1975.

National Science Board. *Science and Engineering Indicators*. Biennial. Washington, DC: National Science Foundation.

Office of Management and Budget, *Social Indicators, 1973*. Washington, DC: U.S. Government Printing Office, 1974.

Rossi, Peter H., Mark W. Lipsey, and Howard E. Freeman. *Evaluation: A Systematic Approach*. Thousand Oaks, CA: Sage Publications, 2003.

Rothman, Jack. *Planning and Organizing for Social Change: Action Principles from Social Science Research*. New York: Columbia University Press, 1974.

Shadish, William, Thomas D. Cook, and Donald T. Campbell. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Belmont, CA: Wadsworth, Cengage Learning, 2002.

Singpurwalla, N. D. "Models in Air Pollution," in *A Guide to Models in Governmental Planning and Operations*, ed. Saul I. Gass and Roger L. Sisson. Washington, DC: U.S. Environmental Protection Agency, 1975.

Smith, H. W. *Strategies of Social Research: The Methodological Imagination*. Englewood Cliffs, NJ: Prentice Hall, 1975.

Trochim, William M. K. *Research Design for Program Evaluation: The Regression-Discontinuity Approach*. Beverly Hills, CA: Sage Publications, 1984.

U.S. Department of Health, Education, and Welfare. *Toward a Social Report*. Ann Arbor: University of Michigan Press, 1970.

U.S. Department of Health and Human Services. *Poverty Guidelines*. Washington, DC, 2017. https://aspe.hhs.gov/poverty-guidelines.

U.S. Department of Labor. *State Economic and Social Indicators*. Washington, DC: U.S. Government Printing Office, 1973.

U.S. Environmental Protection Agency. *Air Quality Index Report*. Washington, DC. www.epa.gov/outdoor-air-quality-data/air-quality-index-report.

The World Bank. Worldwide Governance Indicators, http://info.worldbank.org/governance/wgi/index.aspx#home.

CASE 6.1 MONITORING POLICY OUTCOMES: THE POLITICAL ECONOMY OF TRAFFIC FATALITIES IN EUROPE AND THE UNITED STATES

The 55 mph speed limit was adopted in 1974 as a means to reduce gasoline consumption during the 1973–1974 OPEC oil embargo. Unexpectedly, the policy was followed by a sharp decline in traffic fatalities. Between January 1, 1974, and December 31, 1975, there was a decline of 9,100 fatalities, a 16.8 percent drop. Despite continuing opposition from rural western states and the trucking industry, there was broad political support for the policy. And, clearly, the problem is not unimportant or trivial. Traffic fatalities represent the leading cause of death among persons 35 years of age and younger. Average annual highway deaths since 1974 are equivalent to a fully loaded 767 aircraft crashing with no survivors every third day of the week.

On April 2, 1987, Congress enacted the Surface Transportation and Uniform Relocation Assistance Act of 1987, overriding President Reagan's veto. Provisions of this bill permitted individual states to experiment with speed limits up to 65 mph on rural interstate highways. By July 1988, forty states had raised the speed limit to 60 or 65 mph on 89 percent of rural interstate roads. Senator John C. Danforth, an influential supporter of the 55 mph speed limit, argued against the new policy. The 65 mph speed limit, said Danforth, would save an average of one minute per day per driver, but result in an annual increase of 600 to 1,000 deaths. *The Washington Post* joined the opponents. "The equation is in minutes versus lives. It's not even close. . . . A hundred miles at 55 mph take about 17 minutes longer than at 65. That's the price of those lives. It ought to be the easiest vote the House takes this year." Eight years later, in a November 1995 press release, U.S. Secretary of Transportation Federico Pena reaffirmed the Clinton administration's opposition to the higher speed limits (see Chapter 1, Case 1.2). The National Maximum Speed Law was officially abandoned the same year.

EXHIBIT C6.1

Fatalities and Other Variables in the United States, 1966–2015

	YEAR	YEARS_	FAT	BIMI	IIP	UEMR	EMP	POP
1	1966	1	50894	926	35.7000	3.2	72895	195578
2	1967	2	50724	964	36.4000	3.1	74372	197457
3	1968	3	52725	1016	39.4635	2.9	75920	199399
4	1969	4	53543	1062	40.1779	2.8	77902	201385
5	1970	5	52627	1110	38.7009	4.4	78627	203984
6	1971	6	52542	1179	40.6885	5.3	79120	206827
7	1972	7	54589	1260	45.4161	5.0	81702	209284
8	1973	8	54052	1313	47.6644	4.2	84409	211357
9	1974	9	45196	1281	43.6268	4.9	85936	213342
10	1975	10	44525	1328	44.0037	7.9	84783	215465
11	1976	11	45523	1402	47.5003	7.1	87485	217561
12	1977	12	47878	1467	50.3455	6.3	90546	219758
13	1978	13	50331	1545	53.8850	5.3	94373	222093
14	1979	14	51093	1529	53.7087	5.1	96945	224569
15	1980	15	51091	1527	53.2684	6.9	97722	227255
16	1981	16	49301	1553	51.7685	7.4	98878	229637
17	1982	17	43945	1595	48.6945	9.9	98047	231996
18	1983	18	42589	1653	54.0090	9.9	99347	234284
19	1984	19	44257	1720	56.7436	7.4	103413	236477
20	1985	20	43825	1774	57.5636	7.0	105613	237924
21	1986	21	46087	1836	58.3863	6.9	107860	240133
22	1987	22	46390	1921	62.6829	6.2	110634	242289
23	1988	23	47087	2026	64.4210	5.5	113153	244499
24	1989	24	45582	2096	64.4375	5.2	115453	246819
25	1990	25	44599	2144	63.7612	5.6	116660	249403
26	1991	26	41508	2172	64.2613	7.0	115657	252138
27	1992	27	39250	2247	66.8587	7.8	116396	255039
28	1993	28	40150	2297	69.0884	7.2	118149	257800
29	1994	29	40716	2360	73.8818	6.1	121999	260350
30	1995	30	41817	2423	75.8317	5.6	124766	262755
31	1996	31	41907	2469	80.3307	5.4	125311	265284
32	1997	32	42013	2560	86.9454	4.9	129558	267744
33	1998	33	41471	2619	90.1893	4.5	131463	270299
34	1999	34	41611	2691	94.9777	4.2	133488	279000
35	2000	35	41945	2750	96.0165	4.0	136208	282200
36	2001	36	42196	2797	91.2654	4.7	136933	285000
37	2002	37	43005	2856	93.8334	5.8	136485	287600
38	2003	38	42884	2890	95.5638	6.0	137736	290100

EXHIBIT C6.1 (CONTINUED)

Fatalities and Other Variables in the United States, 1966–2015

39	2004	39	42836	2965	98.9703	5.5	139252	292800
40	2005	40	43510	2989	101.4873	5.1	141730	295500
41	2006	41	42708	3014	103.6229	4.6	144427	298400
42	2007	42	41059	3003	105.7256	4.6	146047	301200
43	2008	43	37423	2977	93.7103	5.8	145362	304100
44	2009	44	33808	2957	90.8895	9.3	139877	306800
45	2010	45	31000	2967	96.4959	9.5	139064	310233
46	2011	46	32479	2950	98.6739	9.2	139869	313232
47	2012	47	33561	2969	101.0382	8.3	142469	316266
48	2013	48	35200	2988	102.9513	8.0	143929	319330
49	2014	49	32675	3026	106.5182	6.6	146305	322423
50	2015	50	35092	3148	104.0462	5.6	148834	326540

Note: YEAR = year; YEARS_ = numbered year in series; FAT = fatalities; BIMI = billions of miles traveled; IIP = Index of Industrial Production; UEMR = unemployment rate; EMP = employment; POP = population. The first column is case number in the SPSS file.

EXHIBIT C6.2

Fatalities per 100,000 Persons, 1970–1976

Year	US	EUREXP	EURCON	UK	FIN	FR	DEN
1970	25.8	21.9	24.3	14.3	24.8	23.5	25
1971	25.4	22.8	24.7	14.8	25.9	25.4	25
1972	26.1	22.8	25.0	14.6	25.2	27.4	24
1973	25.6	22.1	23.5	14.7	23.8	25.6	24
1974	21.2	17.4	21.3	13.0	18.0	22.5	16
1975	20.1	17.6	20.2	11.9	19.0	24.0	17
1976	20.9	18.3	18.8	12.0	20.0	26.0	17

Note: EUREXP = European "experimental" states that adopted a 90 kph (54 mph) speed limit in 1972; EURCON = European "control" states. Other names refer to countries. The fatality rate is expressed as fatalities per 100,000 persons.

Many analysts who have evaluated the effects of the intervention, along with elected officials from the ten northeastern states that retained the 55 mph speed limit until its repeal, affirm that the 1974 law was responsible for the decline in traffic fatalities. However, the evidence suggests that they may have failed to consider rival hypotheses, which—had they been identified and tested—would have resulted in a different explanation of traffic deaths and, consequently, a different policy recommendation.

Here are data on U.S. traffic fatalities and other variables for the period 1966–2015. In addition, fatalities per 100,000 persons in the U.S. and European countries are provided in a separate listing for the years 1970–1976. Use the data for the demonstration exercise. ■

CHAPTER 7

Evaluating Policy Performance

LEARNING OBJECTIVES

By studying this chapter you should be able to:

- Compare and contrast policy-analytic methods of monitoring and evaluation
- List characteristics that distinguish evaluation from other methods
- Define and illustrate criteria for evaluating policy performance
- Contrast evaluation, valuation, and meta-evaluation
- Distinguish values, ethics, and meta-ethics
- Analyze ethical and economic issues related to "living wage" policies

INTRODUCTION

The primary aim of monitoring is to make sound designative claims about the consequences of policies. Monitoring is about facts, while evaluation is about facts *and* values. Monitoring answers the question: Did the policy produce its intended outcome? Evaluation answers a related but distinct question: Of what value is the outcome?

In this chapter we consider the policy-analytic method of evaluation by first reviewing several ways in which ethics and values are important for public policy. Frameworks for thinking about values, ethics, and meta-ethics (meta-ethics is the philosophical study of ethical reasoning) are then introduced. We go on to examine the nature, aims, and functions of evaluation in policy analysis, showing how evaluation as a form of ethical appraisal helps produce information about policy performance.[1] By contrast, monitoring as a type of causal appraisal helps produce information about the effect of policy on outcomes.[2] We then compare and contrast three approaches to evaluation and examine methods and techniques employed in conjunction with these approaches.

ETHICS AND VALUES IN POLICY ANALYSIS

Some analysts believe that policy evaluation should concentrate on explanation and prediction, treating ethics and values as an afterthought. Policy disagreements would diminish or disappear, argue "predictivists," if analysts could predict with greater certainty. There is no point spending much time on differences in basic values, because they are disputed and difficult to resolve.[3]

This view of policy analysis, which is associated with a philosophy of science called logical positivism, makes a sharp separation between facts and values. The acquisition of facts is seen as positive, because facts affirm what *is*, as distinguished speculations about what *ought to be*. Logical positivism[4] was developed in Europe in the 1920s by a group of physicists and philosophers of science called the Vienna Circle. The original Vienna Circle positivists, and later their followers in the social sciences, believed that questions of ethics and morality are not proper subjects of science. Because observation, explanation and prediction are deemed to be the only legitimate aims of science, ethical and value judgments are seen as unscientific, extrascientific, or antiscientific. Science and values should be kept strictly apart. In the social sciences, analysts should confine their work to explaining and predicting policy outcomes, limiting themselves to the observation and analysis of what is rather than what ought to be.

[1]Duncan MacRae Jr., *The Social Function of Social Science* (New Haven, CT: Yale University Press, 1976), rightly observes that policy analysis may be considered a form of applied ethics.

[2]One of the important aims of William R. Shadish, Thomas D. Cook, and Donald T. Campbell, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference* (Belmont, CA: Wadsworth, Cengage Learning, 2002) is to develop and justify a generalized theory of causal inference.

[3]This is a paraphrase from Milton Friedman, *Essays in Positive Economics* (Chicago, IL: University of Chicago Press, 1953), p. 5. Friedman is speaking about "positive economics."

[4]Logical positivism is also called *logical empiricism* and, sometimes, *scientific empiricism*.

Thinking about Values

Positivism was abandoned by philosophers in the mid-1950s. It is no longer regarded as a tenable model of science. The consensus today is that science, which is subject to a great many social, cultural, economic, political, and psychological influences, does not work the way positivists said it does. In this context, analysts need to look critically at the relation (not separation) between facts and values. For this reason, it is important to understand what we mean when we use the terms "values" and "ethics."

The term values may be defined in a comprehensive way, so that it includes any wants, needs, interests, or preferences.[5] Here, values represent "objects" along a broad spectrum of behaviors involving choice, for example, "He prefers guns to butter." or "She wants guns more than butter." In this view, which refers to *values-as-objects* (e.g., the preference for efficiency), values are anything of interest to those who make evaluations. Given this broad conception, policy analysts can be seen to address a sweeping array of values. It is in this broad sense that policy choices involve "the authoritative allocation of values."[6]

A second, more specific conception of values refers to *values-as-criteria*. Here, the concern is not only with values as objects but also with the criteria used to choose values. This second meaning of values incorporates decision rules or criteria employed to choose between two or more values. Whereas a broad definition of values leads to statements such as "The majority of citizens want programs to be more efficient," a more concrete definition specifies the criterion or rule: "The majority of citizens want programs according to the rule: Select the program that is more efficient because it produces the greatest net benefit (i.e., benefits minus costs)."

These two usages of the term value are related to the different contexts—personal, standard, and ideal—in which we understand values.[7] In the personal context, values are *expressed* in the form of preferences, wants, and needs. A value expression is "I prefer public to private schools." By contrast, the standard context involves *statements* about an individual or group that holds certain values, for example, "School busing for racial integration is a good policy in the eyes of middle-class citizens." Finally, the ideal context involves *value judgments* that are not reducible to value expressions or value statements.

Value judgments are exceptional because they justify values such as efficiency expressed with the rule: Select the policy that maximizes net benefits. Accordingly, it is the role of value judgments to justify such decision rules, placing them in their proper context as instances of ethical argumentation (see Chapter 8). For example, "Policy A is preferable to policy B, because it produces the greatest net benefit, thereby increasing the aggregate satisfaction of members of the community—that is, the greatest good for the greatest number." This value judgment goes beyond the description of personal preferences and the values of a particular group, for example, rational choice theorists who may state uncritically that "Maximize net benefits!" is a self-evident moral imperative, when in fact it is a rule requiring ethical

[5]Stephen C. Pepper, *The Sources of Value* (Berkeley: University of California Press, 1958).

[6]This definition, well known to political scientists, is that of David Easton in *The Political System* (Chicago, IL: Alfred Knopf, 1953).

[7]Abraham Kaplan, *The Conduct of Inquiry* (San Francisco, CA: Chandler, 1964), pp. 387–397.

justification. Although we can describe values, we can and should supply grounds for their justification—this is the special role of value judgments.[8]

In addition to differences in the contexts of values, there are other important distinctions.

- *Value versus values.* Because people express values that we can describe and explain, it does not follow that such value expressions refer to things that are of value. The value of something cannot be established on the basis of observation alone, because it is always possible to ask: "Are these values good?" Consider the quip inspired by Oscar Wilde: "An economist is a man who knows the price of everything and the value of nothing."
- *Values versus needs.* Although values may be a source of needs, as when the value of health justifies attempts to satisfy unmet needs for minimum nutritional standards, values are not the same as needs. A perceived deficiency that gives rise to a need (e.g., nutrition as a material need) may be relatively independent of such values as security, order, justice, and equality. Complex relationships between needs and values create conceptual and methodological issues for analysts who conduct "needs assessments," because needs are not the same as values.
- *Values versus norms.* Whereas norms are rules of conduct applicable in specific contexts, values are standards of achievement applicable in many or all contexts. Values of power, respect, rectitude, affection, well-being, wealth, and enlightenment[9] may each result in the establishment of separate sets of norms intended to regulate relationships between analyst and politician, analyst and analyst, analyst and journalist, analyst and client, or analyst and the public at large. Conversely, the norm of public accountability may involve power, respect, well-being, and other values. Although values provide grounds for justifying norms, the more general the norm, the more difficult it is to distinguish it from a value. Codes of conduct for regulating the behavior of planners, policy analysts, and public managers contain norms that are sufficiently general to be virtually indistinguishable from values.[10]
- *Valuation versus evaluation.* Another important distinction is between valuation and evaluation.[11] Evaluation as a general social process involving many forms of appraisal should be distinguished from valuation, a special process involving efforts to establish the grounds on which appraisals are made. Whereas evaluation requires that two or more alternatives be appraised—for example, according to the criterion of efficiency improvement—valuation requires that the criterion itself be justified, for example, in terms of the utilitarian philosophy of the greatest good for the greatest number or in terms of the the deontological principle that promises should be kept, apart from the consequences of doing so. Disputes surrounding the aims and nature of policy and program evaluation often turn on the distinction between valuation and evaluation.[12]

[8]Ibid., pp. 387–389.

[9]See Harold D. Lasswell and Abraham Kaplan, *Power and Society* (New Haven, CT: Yale University Press, 1950).

[10]For illustrations, see Elizabeth Howe and Jerome Kaufman, "Ethics and Professional Practice in Planning and Related Policy Professions," *Policy Studies Journal* 9, 4 (1981): 585–594.

[11]See John Dewey, "Theory of Valuation," *International Encyclopedia of Unified Science* 11, 4 (1939).

[12]See, for example, Ernest House, *Evaluating with Validity* (Beverly Hills, CA: Sage Publications, 1980), and Deborah Stone, "Conclusion: Political Reason," in *Policy Paradox: The Art of Political Decision Making* (New

- *Valuation versus prescription.* Value judgments, which are products of valuation, should not be confused with nonrational emotional appeals, ideological exhortations, or doctrinaire forms of policy advocacy. This confusion is related to the notion that policy analysis must be value-neutral,[13] for so long as valuation is mistakenly believed to be subjective and nonrational, value-neutrality is the only protection against analyses that are seen as unscientific and arbitrary. Many analysts fail to recognize that value judgments are not "commands, prescriptions, or other forms of telling people to do things . . . telling people what they should do is not telling or exhorting them to do it . . . it is purporting to give them sound solutions to their practical problems."[14]

Ethics and Meta-ethics

The term *ethics* refers to the reflective study of moral choice, whereas *meta-ethics* refers to the reflective study of ethics itself. The term ethics is often applied to habitual or customary behavior, as in the "ethics of policymakers." This everyday usage derives etymologically from the Greek and Latin *ethos* and *mores,* which refer to habits or customs. Although it is primarily this everyday usage that has guided the development of ethical codes for public managers, planners, and policy analysts,[15] these efforts at professional standard setting are based on implicit ethical and moral judgments. "Even when 'customary morality' is spoken of, the reference of the term is not merely to the customs as such—in the sense of regular, repeated sequences of behavior—but also to the view, at least implicitly held by the participants, that what they regularly do is in some way right; it is not merely what is done, it is also what is to be done."[16] As can be seen, the term ethics has two different but related meanings. Descriptive ethics, which involves the analysis of descriptions of customary morality, should be contrasted with normative ethics. Normative ethics is concerned with the analysis, evaluation, and development of normative statements that provide guidance to those trying to solve practical problems.

Normative ethics is concerned with the appraisal of criteria for justifying normative statements. The central questions of normative ethics are of the form: What criteria justify normative claims about the rightness, goodness, or justice of public actions? By contrast, meta-ethics is concerned with the nature and meaning of normative claims in general. Questions of meta-ethics are of the form: What (meta-ethical) criteria warrant the choice of (normative ethical) criteria employed to justify claims about the rightness, goodness, or justice of public policies? In this and other cases, the term "meta" signifies something that is "about" or "of" something else. "Metapolicy," for example, is making policy about policies,[17]

York: W. W. Norton, 2002), pp. 384–415.

[13]Stokey and Zeckhauser, *A Primer for Policy Analysis,* pp. 4–5.

[14]Kurt Baier, "What Is Value? An Analysis of the Concept," in *Values and the Future,* ed. Kurt Baier and Nicholas Rescher (New York: Free Press, 1969), p. 53.

[15]For example, Herman Mertins, *Professional Standards and Ethics: A Workbook for Public Administrators* (Washington, DC: American Society for Public Administration, 1979).

[16]Louis A. Gewirth, *Reason and Morality* (Chicago, IL: University of Chicago Press, 1978), p. 67.

[17]Yehezkel Dror, *Public Policymaking Reexamined* (New York: Elsevier, 1968).

and "meta-evaluation" is the evaluation of evaluations, because it develops and applies criteria to evaluate program and policy evaluations.[18]

There has been a growing concern with the role of normative ethics and meta-ethics in public policy analysis. An emphasis on the reflective evaluation of criteria employed to justify normative claims may be found in the general literature of policy analysis.[19] In addition, there have been specialized attempts to examine meta-ethical questions that are ontological (Does ethical knowledge exist?), epistemological (What criteria govern the truth or falsity of ethical claims?), and practical (What is the relation between ethics and moral action?). Michalos, for example, argues that widely accepted but mistaken assumptions about the existence of two mutually exclusive realms of experience ("facts" and "values") leads analysts to the false conclusion that ethical claims are "scientifically incorrigible."[20] Drawing on critical theory and the work of Jurgen Habermas and other members of the Frankfurt School, Dallmayr argues that "policy evaluation requires a critically reflective 'practical discourse' open not only to experts or policy analysts but to the public at large . . . recovery of a fully non-instrumental 'practical' judgment presupposes an evaluation not only of concrete policies but of the status of 'policy' itself."[21]

Distinctions among descriptive ethics, normative ethics, and meta-ethics are particularly applicable to issues of social equity and justice. Issues of distributive justice—that is, disagreements about the appropriate rule for distributing income—are often addressed by describing the consequences of adopting one or more of the following neo-utilitarian decision criteria:

- *Net efficiency improvement.* Maximize total benefits minus total costs.
- *Pareto improvement.* Maximize total benefits minus total costs up to that point where it is not possible to make any person better off without also making another person worse off.
- *Kaldor–Hicks improvement.* Also called *virtual Pareto improvement*, maximize total benefits minus total costs, provided that winners may in principle compensate losers.

At the level of descriptive ethics, the satisfaction of these criteria can be addressed by comparing a measure of inequality such as the Gini index (see Chapter 6), which measures differences in the distribution of household income. By contrast, at the level of normative ethics, the justifiability of these criteria may be addressed by normative theories of justice, for example, John Rawls's theory of justice-as-fairness.[22] In a pre-civil society in which dis-

[18]Thomas D. Cook and Charles Gruder, "Metaevaluation Research," *Evaluation Quarterly* 2 (1978): 5–51. MacRae has developed an important original application of meta-ethics to the normative ethics of policy analysis. Duncan MacRae Jr., *The Social Function of Social Science* (New Haven, CT: Yale University Press, 1976).
[19]For example, MacRae, *The Social Function of Social Science*; Frank Fischer, *Evaluating Public Policy* (Chicago, IL: Nelson-Hall, 1995).
[20]Alex C. Michalos, "Facts, Values, and Rational Decision Making," *Policy Studies Journal* 9, 4 (1981): 544–551.
[21]Fred R. Dallmayr, "Critical Theory and Public Policy," *Policy Studies Journal* 9, 4 (1981): 523. Daneke makes a general case for the introduction of meta-ethics into the curricula of schools of public policy and management. Gregory Daneke, "Beyond Ethical Reductionism in Public Policy Education," in *Ethics, Values, and the Practice of Policy Analysis,* ed. William N. Dunn (Lexington, MA: D.C. Heath, 1982).
[22]John Rawls, *A Theory of Justice* (Cambridge, MA: Harvard University Press, 1968).

crepancies of power, wealth, and privilege are as yet unknown—a society assumed to be under a Rawlsian "veil of ignorance"—citizens would accept a decision criterion different from those used in any actual society. The Rawlsian decision criterion is to maximize the welfare of members of society who are worst off. The justifiability of this criterion can be challenged at a basic, meta-ethical level. In this context, Rawlsian theory has been challenged on meta-ethical grounds that it presupposes an individualistic conception of human nature where normative commitments are no longer products of reasoned social discourse. Instead, normative commitments are

> the contractual composite of arbitrary (even if comprehensible) values individually held and either biologically or socially shaped. . . . The structure of the Rawlsian argument thus corresponds closely to that of instrumental rationality; ends are exogenous, and the exclusive office of thought in the world is to ensure their maximum realization . . . [Rawls's arguments] reduce all thought to the combined operations of formal reason and instrumental prudence in the service of desire.[23]

Standards of Conduct

The standards of conduct or norms of "customary morality" that guide the behavior of analysts are variable. In part, this variability may be attributed to the diverse approaches to teaching of ethics and values in schools of public policy and management.[24] Government policy analysts and planners are themselves divided along methodological and ethical lines.[25] Efforts to reduce this variability have drawn on several sources of standards: social values and norms, including equity, honesty, and fairness, which may guide the actions of policy analysts as well as citizens at large; scientific values and norms, including objectivity, neutrality, and institutionalized self-criticism; professional codes of conduct, including formally stated obligations, duties, and prescriptions; and legal and administrative procedures, including mechanisms for securing informed consent by seeking the formal approval of institutional review boards.

The communication of social values and norms are central to the research and teaching activities of individual scholars, whereas the communication of scientific and professional norms has occurred largely through the curriculum development efforts of the National Association of Schools of Public Affairs and Administration (NASPAA) and through the traditional conference and publications activities of the Policy Studies Organization (PSO), the Association for Public Policy and Management (APPAM), and the American Society for Public Administration (ASPA). In addressing the need for a code of professional conduct, but stopping short of the formal codification of standards, ASPA has published and disseminated a pamphlet titled *Professional Standards and Ethics: A Workbook for Public*

[23]Laurence H. Tribe, "Ways Not to Think about Plastic Trees," in *When Values Conflict,* ed. L. H. Tribe, C. S. Schelling, and J. Voss (Cambridge, MA: Ballinger, 1976), p. 77.

[24]Joel L. Fleishman and Bruce L. Payne, *Ethical Dilemmas and the Education of Policymakers* (Hastings-on-Hudson, NY: Hastings Center, 1980).

[25]Elizabeth Howe and Jerome Kaufman, "The Ethics of Contemporary American Planners," *Journal of the American Planning Association* 45 (1979): 243–255, and "Ethics and Professional Practice in Planning and Related Planning Professions," *Policy Studies Journal* 9, 4 (1981): 585–594.

Administrators. By contrast, legal and administrative procedures to regulate the conduct of research involving human subjects and research integrity generally are products of government initiatives rather than those of professional associations or universities.

DESCRIPTIVE ETHICS, NORMATIVE ETHICS, AND META-ETHICS

Theories of ethics and values may be conveniently classified according to their functions: the description, classification, measurement of ethics and values (descriptive theories); the development and application of criteria to assess ethical behaviors (normative theories); and the development and application of additional criteria to assess normative ethical theories themselves (meta-ethical theories). This classification has important subdivisions—for example, teleological versus deontological normative theories, or cognitivist versus non-cognitivist meta-ethical theories.

Descriptive Value Typologies

The work of Milton Rokeach[26] represents a major synthesis of descriptive theories of values developed since the 1930s. The scope and depth of his concern is evident in the claim that

> the concept of values, more than any other, is the core concept across all the social sciences. It is the main dependent variable in the study of culture, society, and personality, and the main independent variable in the study of social attitudes and behavior. It is difficult for me to conceive of any problem social scientists might be interested in that would not deeply implicate human values.[27]

One of Rokeach's contributions is the development of a basic typology for developing and testing descriptive theories of values. The basic typology has two major dimensions: terminal and instrumental values. Terminal values, which are both personal and social, are beliefs about desirable end-states of existence. Instrumental values are beliefs about desirable modes of conduct (Table 7.1).

Howe and Kaufman have employed a similar typology in their study of the ethics of professional planners.[28] Distinguishing between "ends-oriented" and "means-oriented" ethics regarding the importance of a large number of values reported in existing literature, Howe and Kaufman draw two classes of ethical principles from the 1962 *Code of Professional Responsibility and Rules of Procedure of the American Institute of Planners.* Among the major findings of the study are that professional and personal ethical norms are often inconsistent and that a commitment to certain ends (e.g., social equity) affects ethical judgments about the use of particular means (e.g., leaking information to low-income groups). Although this study is essentially exploratory, it addresses important competing principles:

[26]Milton Rokeach, *The Nature of Human Values* (New York: Free Press, 1973).
[27]Ibid., p. ix.
[28]Howe and Kaufman, "The Ethics of Contemporary American Planners."

TABLE 7.1

Basic Value Typology: Terminal and Instrumental Values

Terminal Value	Instrumental Value
A comfortable life	Ambitious
An exciting life	Broadminded
A sense of accomplishment	Capable
A world at peace	Cheerful
A world of beauty	Clean
Equality	Courageous
Family security	Forgiving
Freedom	Helpful
Happiness	Honest
Inner harmony	Imaginative
Mature love	Independent
National security	Intellectual
Pleasure	Logical
Salvation	Loving
Self-respect	Obedient
Social recognition	Polite
True friendship	Responsible
Wisdom	Self-controlled

Source: Rokeach (1973): Table 2.1, p. 28. Reliability coefficients and parenthetical qualifiers have been omitted.

good ends justify the selection of means[29] versus means should be justified in their own terms.[30]

Developmental Value Typologies

Basic typologies can serve as a basis for examining theories of individual and collective value change and development. For example, Rokeach tests a theory of cognitive inconsistency with data organized according to the typology of terminal and instrumental values.[31] He also identifies a variety of developmental patterns based on age differences, concluding that values change not only during adolescence but also throughout life.[32] These patterns are related to the stage theory of moral development created and tested by Lawrence

[29]See Saul Alinsky, *Rules for Radicals* (New York: Random House, 1971).
[30]See Sissela Bok, *Lying: Moral Choice in Public and Private Life* (New York: Pantheon, 1978).
[31]Rokeach, *Nature of Human Values,* pp. 215–234.
[32]Ibid., p. 81.

Kohlberg.[33] Kohlberg's developmental typology distinguishes three levels in the development of moral reasoning: pre-conventional, conventional, and post-conventional. Because each level has two stages, the typology yields six stages in all: Stage 1 (punishment and obedience orientation); Stage 2 (instrumental relativist orientation); Stage 3 (interpersonal concordance orientation); Stage 4 (law-and-order orientation); Stage 5 (social-contract legalistic orientation); and Stage 6 (universal ethical principle orientation).

Kohlberg's developmental typology is of great potential importance for assessing the role of ethical norms and principles in policy analysis. The six stages are claimed to represent an invariant sequence whereby individuals progress in their *capacity* for moral judgment. Because the sequence is invariant and sequential, developmental transitions can occur only between adjacent stages. Accordingly, individuals cannot be expected to move directly from an egocentric market society (Stage 2) to a society based on universal ethical principles (Stage 6), for example, from a concern with an ethic of income maximization to an ethic of the type outlined by Rawls. Because Kohlberg's theory is based on relatively firm (although imperfect) empirical grounds, it has an important bearing on questions surrounding possibilities for normative ethical discourses in contemporary society: "Any conception of what moral judgment ought to be must rest on an adequate conception of what is. The fact that our conception of the moral 'works' empirically is important for its philosophic adequacy."[34]

Normative Theories

Normative theories of value, or normative ethics, may be divided into three main types: deontological, teleological, and practical. Deontological normative theories claim that certain kinds of actions are inherently right or obligatory (the Greek *deontos* means "of the obligatory"), or right because actions conform to some formal principle. Teleological normative theories, by contrast, hold that certain actions are right because they result in good of valuable ends (the Greek *teleios* means "brought to its end or purpose"). Finally, practical normative theories hold that certain actions are right because they conform to principles, or result in consequences, whose rightness or goodness has been established through reasoned discourses or transactions (the Greek *praktikos* means "to experience, negotiate, or transact") among those who affect and are affected by the creation and application of moral rules.

Deontological normative theories rest on "deontic" concepts and principles, as distinguished from concepts and principles based on the goodness of consequences. Theories of justice such as that offered by Rawls are mainly deontological because they employ formal principles of obligation (a pre-civil social contract executed under a "veil of ignorance") to justify arguments about distributive justice ("justice as fairness") and a just society (a society that maximizes the welfare of those who are worst off). By contrast, teleological theories evaluate actions according to the goodness of their consequences. A prominent form of teleological theory, one that has deeply affected policy analysis through modern welfare

[33] Lawrence Kohlberg, *Stages in the Development of Moral Thought and Action* (New York: Holt, Rinehart, and Winston, 1961).

[34] Lawrence Kohlberg, "From Is to Ought: How to Commit the Naturalistic Fallacy and Get Away with It in the Study of Moral Development," in *Cognitive Development and Epistemology,* ed. T. Mischel (New York: Academic Press, 1971), pp. 151–235. The "naturalistic fallacy" is the fallacy of deriving an "ought" from an "is."

economics, is *utilitarianism*. The classical utilitarian theories of Jeremy Bentham and John Stuart Mill held that right actions are those that promote the maximum or greatest good for everyone. Policy analysts who use cost–benefit analysis are utilitarians to the degree that they base policy recommendations on the criterion of maximizing net income benefits, a criterion presumed to reflect the aggregate satisfaction experienced by members of society.

The main properties of teleological theories, including modern neo-utilitarian theories, are most easily visualized by comparing them with deontological theories.[35]

- Teleological theories justify actions because of their consequences (e.g., maximization of net income benefits), whereas deontological theories justify actions because they conform to some principle (e.g., procedural fairness) or because they are deemed to be inherently right (e.g., truth-telling).
- Teleological theories set forth conditional obligations (e.g., freedom might be compromised in the interests of national security), whereas deontological theories set forth absolute obligations (e.g., freedom is a generic right that cannot be justified in terms of any higher principle).
- Teleological theories advance material or substantive criteria (e.g., pleasure, happiness, or satisfaction), whereas deontological theories advance formal or relational criteria (e.g., social equity or administrative impartiality).
- Teleological theories supply aggregate criteria (e.g., social welfare as the total satisfaction experienced by members of a community), whereas deontological theories supply distributive criteria regulating the allocation of goods (e.g., educational opportunities or a living wage) or evils (e.g., exposure to toxic wastes and carcinogens).

Practical normative theories stress reflective moral action as a process for discovering the criteria on which values and ethics may be known. Practical normative theories have stressed reasoned moral discourse as a process for creating and evaluating ethical as well as scientific knowledge.[36] Sometimes called "good reasons" theories, these theories are "practical" to the extent that actions are deemed right or valuable because they conform to principles, or result in consequences, that have been established on the basis of reasoned transactions (Greek *praktikos:* "to experience, negotiate, or transact") among persons who affect and are affected by the development and application of moral rules.

Meta-ethical Theories

The function of normative ethical theories in policy analysis is to answer the question: According to what criteria can we determine whether public actions are right or wrong? Answers to this question, as we saw earlier, may be based on one or more types of normative

[35]See Louis A. Gewirth, *Reason and Morality* (Chicago, IL: University of Chicago Press, 1978).

[36]See P. W. Taylor, *Normative Discourse* (Englewood Cliffs, NJ: Prentice Hall, 1971); Stephen Toulmin, *The Place of Reason in Ethics* (Oxford: Oxford University Press, 1950); Stephen Toulmin, *The Uses of Argument* (Cambridge: Cambridge University Press, 1958); Kurt Baier, *The Moral Point of View* (Ithaca, NY: Cornell University Press, 1965); and Frank Fischer, *Politics, Values, and Public Policy* (Boulder, CO: Westview Press, 1980).

ethical criteria: teleological, deontological, practical. The function of meta-ethical theories, by contrast, is to answer questions about normative ethical claims themselves: Can we determine the truth and falsity of normative ethical claims? Does normative ethics produce a kind of knowledge and, if so, what kind of knowledge is it? If normative ethics is not capable of being true and false, what kinds of non-cognitive results are produced?

Meta-ethical theories differ in terms of their assumptions about the epistemological status of normative ethical theories—for example, so-called "cognitivist" meta-ethics affirms that normative ethical theories are capable of being true or false and, therefore, that normative ethics constitutes a kind of knowledge. This is denied by "non-cognitivist" meta-ethical theories. Meta-ethical theories are associated with normative ethical theories. Thus, for example, the growth of logical positivism in the social sciences has contributed to non-cognitivist meta-ethical doctrines that, in turn, have resulted in the devaluation of normative ethical discourse and an inability to recognize that putative empirical theories (welfare economics) and analytic routines (cost–benefit analysis) are based on controversial ethical premises. In this case, a particular meta-ethical doctrine (noncognitivism) has exerted a direct and logically constraining effect on normative ethics by preempting opportunities for reflective normative discourse and the development of new ethical knowledge.

EVALUATION IN POLICY ANALYSIS

Descriptive, normative, and meta-ethical theories provide a foundation for evaluation in policy analysis, where evaluation refers to the production of information about the value or worth of policy outcomes. When policy outcomes do in fact have value, it is because they contribute to goals and objectives. In this case, we say that a policy or program has attained some significant level of performance.

The Nature of Evaluation

The main feature of evaluation is that it results in claims that are evaluative in character. Here the main question is not one of facts (*Does something exist?*) or of action (*What should be done?*) but one of values (*Of what worth is it?*). Evaluation therefore has several characteristics that distinguish it from other policy-analytic methods:

1. *Value focus.* Evaluation, as contrasted with monitoring, focuses on judgments regarding the desirability or value of policies and programs. Evaluation is primarily an effort to determine the worth or social utility of a policy or program, and not simply an effort to collect information about the anticipated and unanticipated outcomes of policy actions. As the appropriateness of policy goals and objectives can always be questioned, evaluation includes the evaluation of goals and objectives themselves.
2. *Fact-value interdependence.* Evaluation depends as much on "facts" as it does on "values." To claim that a particular policy or program has attained a high (or low) level of performance requires, not only that policy outcomes are valuable to some individual, group, or society as a whole, but it also requires that policy outcomes are

actually a consequence of actions undertaken to resolve a problem. Hence, monitoring is a prerequisite of evaluation.

3. *Present and past orientation.* Evaluative claims, as contrasted with prescriptive claims, are oriented toward present and past outcomes, rather than future ones. Evaluation is retrospective and occurs after actions have been taken (*ex post*). Prescription, while also involving value premises, is prospective and occurs before actions have been taken (*ex ante*).
4. *Value duality.* The values underlying evaluative claims have a dual quality, because they may be regarded as ends and means. Evaluation is similar to prescription insofar as a given value (e.g., health) may be regarded as intrinsic (valuable in itself) as well as extrinsic (desirable because it leads to some other end). Values are often arranged in a hierarchy that reflects the relative importance and interdependency of goals and objectives.

Functions of Evaluation

Evaluation performs several main functions in policy analysis. First, and most important, evaluation provides reliable and valid information about *policy performance,* that is, the extent to which needs, values, and opportunities have been realized through public action. In this respect, evaluation reveals the extent to which particular goals (e.g., improved health) and objectives (e.g., a 20 percent reduction of chronic diseases by 1990) have been attained.

Second, evaluation contributes to the *clarification* and *critique* of values that underlie the selection of goals and objectives. Values are clarified by defining and operationalizing goals and objectives. Values are also critiqued by systematically questioning the appropriateness of goals and objectives in relation to the problem being addressed. In questioning the appropriateness of goals and objectives, analysts may examine alternative sources of values (e.g., public officials, vested interests, client groups) as well as their grounds in different forms of rationality (technical, economic, legal, social, substantive).

Third, evaluation may contribute to the application of other policy-analytic methods, including *problem structuring* and *prescription.* Information about inadequate policy performance may contribute to the restructuring of policy problems, for example, by showing that goals and objectives should be redefined. Evaluation can also contribute to the definition of new or revised policy alternatives by showing that a previously favored policy alternative should be abandoned and replaced with another one.

Criteria for Policy Evaluation

In producing information about policy performance, analysts use different types of criteria to evaluate policy outcomes. These types of criteria have already been discussed in relation to policy recommendations (Chapter 5). The main difference between criteria for evaluation and criteria for prescription is the time at which criteria are applied. Criteria for evaluation are applied retrospectively (*ex post*), whereas criteria for recommendation are applied prospectively (*ex ante*). These criteria are summarized in Table 7.2.

TABLE 7.2

Criteria for Evaluation

Type of Criterion	Question	Illustrative Criteria
Effectiveness	Has a valued outcome been achieved?	Units of service
Efficiency	How much effort was required to achieve a valued outcome?	Unit cost Net benefits Cost–benefit ratio
Adequacy	To what extent does the achievement of a valued outcome resolve the problem?	Fixed costs (Type I problem) Fixed effectiveness (Type II problem)
Equity	Are costs and benefits distributed equitably among different groups?	Pareto criterion Kaldor–Hicks criterion Rawls's criterion
Responsiveness	Do policy outcomes satisfy the needs, preferences, or values of particular groups?	Consistency with citizen surveys
Appropriateness	Are desired outcomes (objectives) actually worthy or valuable?	Public programs should be equitable as well as efficient

Note: See Chapter 5 for extended descriptions of criteria.

APPROACHES TO EVALUATION

Evaluation, as we saw earlier, has two interrelated aspects: the use of various methods to monitor the outcomes of public policies and programs and the application of some set of values to determine the worth of these outcomes to some person, group, or society as a whole. Observe that these two interrelated aspects point to the presence of factual and value premises in any evaluative claim. Yet many activities described as "evaluation" in policy analysis are essentially *nonevaluative*—that is, they are primarily concerned with the production of designative (factual) claims rather than evaluative ones. In fact, each of the four approaches to monitoring described in Chapter 6 is frequently mislabeled as approaches to "evaluation research" or "policy evaluation."[37]

Given the present lack of clarity about the meaning of evaluation in policy analysis, it is essential to distinguish among several different approaches to policy evaluation: pseudo-evaluation, formal evaluation, and decision-theoretic evaluation. These approaches and their aims, assumptions, and major forms are illustrated in Table 7.3.

[37]See, for example, the description of "evaluation" research in Marcia Guttentag and Elmer Struening, ed., *Handbook of Evaluation Research,* 2 vols. (Beverly Hills, CA: Sage Publications, 1975), and Leonard Rutman, ed., *Evaluation Research Methods: A Basic Guide* (Beverly Hills, CA: Sage Publications, 1977).

TABLE 7.3

Three Approaches to Evaluation

Approach	Aims	Assumptions	Major Forms
Pseudo-evaluation	Use descriptive methods to produce reliable and valid information about policy outcomes	Measures of worth or value are self-evident or uncontroversial	Social experimentation Social systems accounting Social auditing Research and practice synthesis
Formal evaluation	Use descriptive methods to produce reliable and valid information about policy outcomes that have been formally announced as policy-program objectives	Formally announced goals and objectives of policymakers and administrators are appropriate measures of worth or value	Developmental evaluation Experimental evaluation Retrospective process evaluation Retrospective outcome evaluation
Decision-theoretic evaluation	Use descriptive methods to produce reliable and valid information about policy outcomes that are explicitly valued by multiple stakeholders	Formally announced as well as latent goals and objectives of stakeholders are appropriate measures of worth or value	Evaluability assessment Multiattribute utility analysis

Pseudo-evaluation

Pseudo-evaluation is an approach that uses descriptive methods to produce reliable and valid information about policy outcomes, without attempting to question the worth or value of these outcomes to persons, groups, or society as a whole. The major assumption of pseudo-evaluation is that measures of worth or value are self-evident or uncontroversial.

In pseudo-evaluation, the analyst typically uses a variety of methods (quasi-experimental design, questionnaires, random sampling, statistical techniques) to explain variations in policy outcomes in terms of policy input and process variables. Yet any given policy outcome (e.g., number of employed trainees, units of medical services delivered, net income benefits produced) is taken for granted as an appropriate objective. Major forms of pseudo-evaluation include the approaches to monitoring discussed in Chapter 6: social experimentation, social systems accounting, social auditing, and research and practice synthesis.

Formal Evaluation

Formal evaluation is an approach that uses descriptive methods to produce reliable and valid information about policy outcomes but evaluates such outcomes on the basis of policy-program objectives that have been formally announced by policymakers and program administrators. The major assumption of formal evaluation is that formally announced goals and objectives are appropriate measures of the worth or value of policies and programs.

In formal evaluation, the analyst uses the same kinds of methods as those employed in pseudo-evaluation and the aim is identical: to produce reliable and valid information about variations in policy outputs and impacts that may be traced to policy inputs and processes. The difference, however, is that formal evaluations use legislation, program documents, and interviews with policymakers and administrators to identify, define, and specify formal goals and objectives. The appropriateness of these formally announced goals and objectives is not questioned. In formal evaluations, the types of evaluative criteria most frequently used are those of effectiveness and efficiency.

One of the major types of formal evaluation is the *summative evaluation,* which involves an effort to monitor the accomplishment of formal goals and objectives after a policy or program has been in place for some period of time. Summative evaluations are designed to appraise products of stable and well-established public policies and programs. By contrast, a *formative evaluation* involves efforts to continuously monitor the accomplishment of formal goals and objectives. The differences between summative and formative evaluation should not be overemphasized, however, because the main distinguishing characteristic of formative evaluation is the number of points in time at which policy outcomes are monitored. Hence, the difference between summative and formative evaluation is mainly one of degree.

Formal evaluations may be summative or formative, but they may also involve direct or indirect controls over policy inputs and processes. In the former case, evaluators can directly manipulate expenditure levels, the mix of programs, or the characteristics of target groups—that is, the evaluation may have one or more characteristics of social experimentation as an approach to monitoring (Chapter 6). In the case of indirect controls, policy inputs and processes cannot be directly manipulated; rather they must be analyzed retrospectively on the basis of actions that have already occurred. Four types of formal evaluation—each based on a different orientation toward the policy process (summative versus formative) and type of control over action (direct versus indirect)—are illustrated in Table 7.4.

Varieties of Formal Evaluation

Developmental evaluation refers to evaluation activities that are explicitly designed to serve the day-to-day needs of program staff. Developmental evaluation is useful "for alerting the staff to incipient weaknesses or unintended failures of a program and for insuring proper operation by those responsible for its operation."[38] Developmental evaluation, which involves some measure of direct control over policy actions, has been used in a wide variety of situations in the public and private sectors. Thus, for example, businesses frequently use developmental evaluations to distribute, test, and recall new products. In the public sector,

[38]Peter H. Rossi and Sonia R. Wright, "Evaluation Research: An Assessment of Theory, Practice, and Politics," *Evaluation Quarterly* 1, 1 (February 1977): 21.

TABLE 7.4

Types of Formal Evaluation

Control over Policy Actions	Orientation toward Policy Process: *Formative*	*Summative*
Direct	Developmental evaluation	Experimental evaluation
Indirect	Retrospective process evaluation	Retrospective outcome evaluation

developmental evaluations have been used to test new teaching methods and materials in public education programs, such as *Sesame Street* and *The Electric Company.* Such programs are systematically monitored and evaluated by showing them to audiences composed of children within specified age limits. Subsequently, they are "revised many times on the basis of systematic observations of which program features achieved attention and on the basis of interviews with the children after viewing the program."[39] Developmental evaluations, as they are both formative and involve direct controls, can be used to adapt immediately to new experience acquired through systematic manipulations of input and process variables.

Retrospective process evaluation involves the monitoring and evaluation of programs after they have been in place for some time. Retrospective process evaluation, which often focuses on problems and bottlenecks encountered in the implementation of policies and programs, does not permit the direct manipulation of inputs (e.g., expenditures) and processes (e.g., alternative delivery systems). Rather it relies on *ex post facto* (retrospective) descriptions of ongoing program activities, which are subsequently related to outputs and impacts. Retrospective process evaluation requires a well-established internal reporting system that permits the continuous generation of program-related information (e.g., the number of target groups served, the types of services provided, and the characteristics of personnel employed to staff programs). Management information systems in public agencies sometimes permit retrospective process evaluations, provided they contain information on processes as well as outcomes.

Title I of the Elementary and Secondary Education Act (1965) was subjected to a form of retrospective process evaluation by the Department of Education, but with disappointing results. Title I provided funds to local school systems in proportion to the number of pupils from poor or deprived families. Local school districts submitted inadequate and marginally useful information, thus making it impossible to evaluate and implement programs concurrently. Retrospective process evaluations presuppose a reliable and valid information system, which is often difficult to establish.

Experimental evaluation involves the monitoring and evaluation of outcomes under conditions of direct controls over policy inputs and processes. The ideal of experimental evaluation has generally been the "controlled scientific experiment," where all factors that might influence policy outcomes except one—that is, a particular input or process variable—are controlled, held constant, or treated as plausible rival hypotheses. Experimental and

[39]Ibid., p. 22.

quasi-experimental evaluations include the New Jersey–Pennsylvania Income Maintenance Experiment, the California Group Therapy–Criminal Recidivism Experiment, the Kansas City Preventive Patrol Experiment, Project Follow Through, the Supported Work Demonstration Project, and various experiments in educational performance contracting.

Experimental evaluations must meet rather severe requirements before they can be carried out:[40] (1) a clearly defined and directly manipulable set of "treatment" variables that are specified in operational terms; (2) an evaluation strategy that permits maximum generalizability of conclusions about performance to many similar target groups or settings (*external validity*); (3) an evaluation strategy that permits minimum error in interpreting policy performance as the actual result of manipulated policy inputs and processes (*internal validity*); and (4) a monitoring system that produces reliable data on complex interrelationships among preconditions, unforeseen events, inputs, processes, outputs, impacts, and side effects and spillovers (see Figure 6.2). As these demanding methodological requirements are rarely met, experimental evaluations typically fall short of the "true" controlled experiment, and are referred to as "quasi-experimental."

Retrospective outcome evaluations also involve the monitoring and evaluation of outcomes but with no direct control over manipulable policy inputs and processes.[41] At best, controls are indirect or statistical—that is, the evaluator attempts to isolate the effects of many different factors by using quantitative methods. In general, there are two main variants of retrospective process evaluation: cross-sectional and longitudinal studies. *Longitudinal studies* are those that evaluate changes in the outcomes of one, several, or many programs at two or more points in time. Many longitudinal studies have been carried out in the area of family planning, where fertility rates and changes in the acceptance of contraceptive devices are monitored and evaluated over reasonably long periods of time (5 to 20 years). *Cross-sectional studies,* by contrast, seek to monitor and evaluate multiple programs at one point in time. The goal of the cross-sectional study is to discover whether the outputs and impacts of various programs are significantly different from one another, and if so, what particular actions, preconditions, or unforeseen events might explain the difference.

Two prominent examples of retrospective outcome evaluations that are cross-sectional in nature are Project Head Start, a program designed to provide compensatory education to preschool children, and the Coleman Report. Indeed,

> almost every evaluation of the national compensatory education programs started during the middle 1960s has been based on cross-sectional data. Pupils enrolled in compensatory education programs were contrasted to those who were not, holding constant statistically such sources of competing explanations as family backgrounds, ethnicity, region, city size, and so on.[42]

[40]Walter Williams, *Social Policy Research and Analysis* (New York: Elsevier, 1971), p. 93.
[41]Note that three of the approaches to monitoring discussed in Chapter 8—that is, social systems accounting, social auditing, and research and practice synthesis—may also be considered as forms of retrospective outcome evaluation. Similarly, cost–benefit and cost–effectiveness analysis (Chapter 5) may be viewed as particular forms of retrospective outcome evaluation.
[42]Rossi and Wright, "Evaluation Research," p. 27.

Decision-Theoretic Evaluation

Decision-theoretic evaluation is an approach that uses descriptive methods to produce reliable and valid information about policy outcomes that are explicitly valued by multiple stakeholders. The key difference between decision-theoretic evaluation, on the one hand, and pseudo-evaluation and formal evaluation on the other, is that decision-theoretic evaluation attempts to surface and make explicit the latent as well as manifest goals and objectives of stakeholders. This means that formally announced goals and objectives of policymakers and administrators are but one source of values, because all parties who have a stake in the formulation and implementation of a policy (e.g., middle- and lower-level staff, personnel in other agencies, client groups) are involved in generating the goals and objectives against which performance is measured.

Decision-theoretic evaluation is a way to overcome several deficiencies of pseudo-evaluation and formal evaluation:[43]

1. *Underutilization and non-utilization of performance information.* Much of the information generated through evaluations is underutilized or never used to improve policymaking. In part, this is because evaluations are not sufficiently responsive to the goals and objectives of parties who have a stake in formulating and implementing policies and programs.
2. *Ambiguity of performance goals.* Many goals of public policies and programs are vague. This means that the same general goal—for example, to improve health or encourage better conservation of energy—can and do result in specific objectives that conflict with one another. This is evident when we consider that the same goal (e.g., improved health) may be operationalized in terms of at least six types of evaluation criteria: effectiveness, efficiency, adequacy, equity, responsiveness, and appropriateness. One of the purposes of decision-theoretic evaluation is to reduce the ambiguity of goals and make conflicting objectives explicit.
3. *Multiple conflicting objectives.* The goals and objectives of public policies and programs cannot be satisfactorily established by focusing on the values of one or several parties (e.g., Congress, a dominant client group, or a head administrator). In fact, multiple stakeholders with conflicting goals and objectives are present in most situations requiring evaluation. Decision-theoretic evaluation attempts to identify these multiple stakeholders and surface their goals and objectives.

One of the main purposes of decision-theoretic evaluation is to link information about policy outcomes with the values of multiple stakeholders. The assumption of decision-theoretic evaluation is that formally announced as well as latent goals and objectives of stakeholders are appropriate measures of the worth or value of policies and programs. The two major forms of decision-theoretic evaluation are evaluability assessment and

[43]For accounts of these and other deficiencies, see Carol H. Weiss, *Evaluation Research* (Englewood Cliffs, NJ: Prentice Hall, 1972); Ward Edwards, Marcia Guttentag, and Kurt Snapper, "A Decision-Theoretic Approach to Evaluation Research," in *Handbook of Evaluation Research,* ed. Guttentag and Struening, pp. 139–181; and Martin Rein and Sheldon H. White, "Policy Research: Belief and Doubt," *Policy Analysis* 3, 2 (1977): 239–272.

multiattribute utility analysis, both of which attempt to link information about policy outcomes with the values of multiple stakeholders.

Evaluability assessment is a set of procedures designed to analyze the decision-making system that is supposed to benefit from performance information and to clarify the goals, objectives, and assumptions against which performance is to be measured.[44] The basic question in evaluability assessment is whether a policy or program can be evaluated at all. For a policy or program to be evaluable, at least three conditions must be present: a clearly articulated policy or program; clearly specified goals and/or consequences; and a set of explicit assumptions that link policy actions to goals and/or consequences.[45] In conducting an evaluability assessment, analysts follow a series of steps that clarify a policy or program from the standpoint of the intended users of performance information and the evaluators themselves:[46]

1. *Policy-program specification.* What federal, state, or local activities and what goals and objectives constitute the program?
2. *Collection of policy-program information.* What information must be collected to define policy-program objectives, activities, and underlying assumptions?
3. *Policy-program modeling.* What model best describes the program and its related objectives and activities, from the point of view of intended users of performance information? What causal assumptions link actions to outcomes?
4. *Policy-program evaluability assessment.* Is the policy-program model sufficiently unambiguous to make the evaluation useful? What types of evaluation studies would be most useful?
5. *Feedback of evaluability assessment to users.* After presenting conclusions about policy-program evaluability to intended users, what appear to be the next steps that should (or should not) be taken to evaluate policy performance?

A second form of decision-theoretic evaluation is *multiattribute utility analysis.*[47] Multiattribute utility analysis is a set of procedures designed to elicit from multiple stakeholders subjective judgments about the probability of occurrence and value of policy outcomes. The strengths of multiattribute utility analysis are that it explicitly surfaces the value judgments of multiple stakeholders; it recognizes the presence of multiple conflicting objectives in policy-program evaluation; and it produces performance information that is more usable from the standpoint of intended users. The steps in conducting a multiattribute utility analysis are the following:

1. *Stakeholder identification.* Identify the parties who affect and are affected by a policy or program. Each of these stakeholders will have goals and objectives that they wish to maximize.

[44]On evaluability assessment, see Joseph S. Wholey and others, "Evaluation: When Is It Really Needed?" *Evaluation* 2, 2 (1975): 89–94; and Wholey, "Evaluability Assessment," *Evaluation Research Methods*, ed. Rutman, pp. 41–56.

[45]See Rutman, *Evaluation Research Methods*, p. 18.

[46]Wholey, "Evaluability Assessment," pp. 43–55.

[47]See Edwards, Guttentag, and Snapper, "A Decision-Theoretic Approach," pp. 148–159.

2. *Specification of relevant decision issues.* Specify the courses of action or inaction about which there is a disagreement among stakeholders. In the simplest case, there will be two courses of action: the status quo and some new initiative.
3. *Specification of policy outcomes.* Specify the range of consequences that may follow each course of action. Outcomes may be arranged in a hierarchy where one action has several consequences, each of which itself has further consequences. A hierarchy of outcomes is similar to an objectives tree (Chapter 5), except that outcomes are not objectives until they have been explicitly valued.
4. *Identification of attributes of outcomes.* Here the task is to identify all relevant attributes that make outcomes worthy or valuable. For example, each outcome may have different types of benefits and costs to different target groups and beneficiaries.
5. *Attribute ranking.* Rank each value attribute in order of importance. For example, if increased family income is an outcome of a poverty program, this outcome may have several value attributes: sense of family well-being; greater nutritional intake; and more disposable income for health care. These attributes should be ranked in order of their relative importance to one another.
6. *Attribute scaling.* Scale attributes should be ranked in order of importance. To do so, arbitrarily assign the least important attribute a value of 10. Proceed to the next most important attribute, answering the question: How many times more important is this attribute than the next-least-important one? Continue this scaling procedure until the most important attribute has been compared with all others. Note that the most important attribute may have a scale value 10, 20, 30, or more times that of the least important attribute.
7. *Scale standardization.* The attributes that have been scaled will have different maximum values for different stakeholders. For example, one stakeholder may give attribute *A,* a value of 60; attribute *B,* a value of 30; and attribute *C,* a value of 10. Another stakeholder, however, may give these same attributes values of 120, 60, and 10. To standardize these scales, sum all original values for each scale, divide each original value by its respective sum, and multiply by 100. This results in separate scales whose component values sum to 100.
8. *Outcome measurement.* Measure the degree that each outcome is likely to result in the attainment of each attribute. The maximum probability should be given a value of 100; the minimum probability should be given a value of 0 (i.e., there is no chance that the outcome will result in the attainment of the attribute).
9. *Utility calculation.* Calculate the utility (value) of each outcome by using the formula:

$$U_i = \Sigma w_i u_{ij}$$

where

U_i = the aggregate utility (value) of the ith outcome
w_i = the standardized scale value of the jth attribute
u_{ij} = the probability of occurrence of the ith outcome on the jth attribute

10. *Evaluation and presentation.* Specify the policy outcome with the greatest overall performance, and present this information to relevant decision-makers.

The strength of multiattribute utility analysis is that it enables analysts to deal systematically with conflicting objectives of multiple stakeholders. This is possible, however, only when the steps just described are carried out as part of a *group process* involving relevant stakeholders. Hence, the essential requirement of multiattribute utility analysis is that stakeholders who affect and are affected by a policy or program are active participants in the evaluation of policy performance.

METHODS FOR EVALUATION

A number of methods and techniques can assist analysts in evaluating policy performance. Nearly all of these techniques, however, may also be used in conjunction with other policy-analytic methods, including problem structuring, forecasting, prescription, and monitoring. Thus, for example, argumentation analysis (Chapter 8) may be used to surface assumptions about expected relationships between policy actions and objectives. Cross-impact analysis (Chapter 4) may prove useful in identifying unanticipated policy outcomes that work against the achievement of policy-program objectives. Similarly, discounting (Chapter 5) may be as relevant to policy-program evaluation as it is to prescription, given that cost–benefit and cost–effectiveness analysis may be used retrospectively (*ex post*) as well as prospectively (*ex ante*). Finally, techniques that range from graphic displays and index numbers to control-series analysis (Chapter 6) may be essential for monitoring policy outcomes as a prelude to their evaluation.

The fact that various techniques may be used with more than one policy-analytic method points to the interdependence of problem structuring, forecasting, prescription, monitoring, and evaluation in policy analysis. Many methods and techniques are relevant to pseudo-evaluation, formal evaluation, and decision-theoretic evaluation (Table 7.5).

Only one of the techniques listed in Table 7.5 has not already been described. *User-survey analysis* is a set of procedures for collecting information about the evaluability of a policy or program from intended users and other stakeholders.[48] User surveys are central to the conduct of evaluability assessments and other forms of decision-theoretic evaluation. The major instrument for collecting information is an interview protocol with a series of open-ended questions. Responses to these questions provide the information required to complete the several steps in an evaluability assessment previously described: policy-program specification; policy-program modeling; policy-program assessment; and presentation of evaluability assessment to users. A sample interview protocol for a user survey analysis is presented in Table 7.6.

[48]See Wholey, "Evaluability Assessment," pp. 44–49.

TABLE 7.5

Techniques for Evaluation by Three Approaches

Approach	Technique
Pseudo-evaluation	Graphic displays
	Tabular displays
	Index numbers
	Interrupted time-series analysis
	Control-series analysis
	Regression-discontinuity analysis
Formal evaluation	Objectives mapping
	Value clarification
	Value critique
	Constraint mapping
	Cross-impact analysis
	Discounting
Decision-theoretic evaluation	Brainstorming
	Argumentation analysis
	Policy Delphi
	User-survey analysis

TABLE 7.6

Interview Protocol for User-Survey Analysis

Step in Evaluability Assessment	Questions
Policy-program specification	1. What are the objectives of the policy or program?
	2. What would be acceptable evidence of the achievement of policy-program objectives?(1)
Policy-program modeling	3. What policy actions (for example, resources, guidelines, staff activities) are available to achieve objectives?(2)
	4. Why will action *A* lead to objective *O*?(2)
Policy-program evaluability assessment	5. What do various stakeholders (for example, Congress, OMB, state auditor general, mayor's office) expect of the program in terms of performance? Are these expectations consistent?
	6. What is the most serious obstacle to achieving objectives?
Feedback of evaluability assessment to users	7. What performance information do you need on the job? Why?
	8. Are present sources of performance information adequate? Why? Why not?
	9. What is the most important source of performance information you will need in the next year?
	10. What key issues should any evaluation address?

Notes:
(1) Answers to this question yield operational measures of objectives.
(2) Answers to these questions yield causal assumptions about the relation between actions and objectives.

Source: Adapted from Wholey (1977): Figure 3, p. 48.

CHAPTER SUMMARY

This chapter has provided an overview of the process of evaluation, contrasting monitoring and evaluation and their very different functions in policy analysis. After distinguishing valuation, evaluation and meta-evaluation, the chapter identifies the role of values, ethics, and meta-ethics in policy analysis. In analyzing economic and ethical issues of social equity and justice, the chapter concludes with cases in redistributional ethics and the morality of economic decisions regarding the living wage.

REVIEW QUESTIONS

1. Compare and contrast policy-analytic methods of evaluation and prescription in terms of time and the types of claims produced by each.
2. Many policy-program evaluations fail to recognize the latent purposes of evaluation, including a desire (a) to make programs look good by focusing on their surface characteristics ("eyewash"); (b) to cover up program failures ("white-wash"); (c) to destroy a program ("submarine"); (d) to engage in evaluation merely as a ritual that must be practiced to receive funding ("posture"); and (e) to postpone attempts to resolve problems ("postponement"). See Edward A. Suchman, "Action for What? A Critique of Evaluative Research," in *Evaluating Action Programs*, ed. Carol H. Weiss (Boston: Allyn and Bacon, 1972), p. 81. What problem does this raise for defining objectives against which performance is evaluated?
3. Compare and contrast formative and summative evaluation. Which of these two types of evaluation provides performance information that is likely to be most useful to policymakers? Why?
4. What are the strengths and limitations of cost–benefit analysis as an approach to evaluation? (see Chapter 5). In your answer refer to contrasts among pseudo-evaluation, formal evaluation, and decision-theoretic evaluation.
5. Select a policy or program that you would like to evaluate. Compare and contrast evaluability assessment and multiattribute utility analysis as techniques of evaluation, indicating which of the two procedures is likely to yield the most reliable, valid, and useful results.

DEMONSTRATION EXERCISE

Issues of a "living wage" turn on questions of poverty, inequality, income distribution, and economic efficiency. Inequalities of many kinds have increased in the past 35 and more years, irrespective of whether Democrats or Republicans were in power in the United States, although conservatives continue to refer to American society as an "opportunity society" and liberals as an "inequality society."

BIBLIOGRAPHY

Bok, Sissela. *Lying: Moral Choice in Public and Private Life*. New York: Pantheon Books, 1978.

Dasgupta, Partha. "Valuation and Evaluation: Measuring the Quality of Life and Evaluating Policy." Background paper prepared for the World Bank's annual *World Development Report* (2000).

Hausman, Daniel M., and Michael S. McPherson. *Economic Analysis, Moral Philosophy and Public Policy*. Cambridge: Cambridge University Press, 2006.

MacRae, Duncan, Jr. *The Social Function of Social Science*. New Haven, CT: Yale University Press, 1976.

Nozick, Robert. *Anarchy, State, and Utopia*. New York: Basic Books, 1974.
Rawls, John. *A Theory of Justice*. Cambridge, MA: Harvard University Press, 1971.
Rokeach, Milton. *The Nature of Human Values*. New York: Free Press, 1973.
Rossi, Peter H., Mark W. Lipsey, and Howard E. Freeman. *Evaluation: A Systematic Approach*. Thousand Oaks, CA: Sage Publications, 2003.
Sen, Amartya, *Inequality Reexamined.* Oxford: Clarendon Press, 1992.
Stone, Deborah, *Policy Paradox: The Art of Political Decision Making*. Rev. ed. New York: W. W. Norton, 2002.
Weiss, Carol H. *Evaluation.* Upper Saddle River, NJ: Prentice Hall, 1998.
Wholey, Joseph S. "Evaluability Assessment," in *Evaluation Research Methods: A Basic Guide*, ed. Leonard Rutman. Beverly Hills, CA: Sage Publications, 1977.

CASE 7.1 THE ETHICS OF CAKE CUTTING

Attempts to define inequality often use the analogy of a cake that can be cut into slices according to different rules.[49] These rules, along with their underlying justifications, deal with the distribution of income, wealth, educational opportunity, and health care. Some of these rules are the basis of theories of justice put forth by well-known political and ethical theorists including John Rawls[50] and Robert Nozick.[51] Rawls has given us a theory of justice that is based primarily on what he sees as a fair distribution of social goods. For Rawls, the fundamental rule of fairness is "allocate social goods so that those worse off are made better off." Nozick, by contrast, has a theory of justice based on what he sees as a fair process for achieving social goods. Here the rule is "allocate social goods in accordance with the rule that everyone has an equal opportunity to acquire them." Both theorists have had a significant influence on the way many people think about values, ethics, and public policy.

1. Divide into three groups. Each group will use different rules for distributing a cake among members of the class. Group A will use three rules for distributing a social good in accordance with *who benefits*. By contrast, group B will use two rules based on *what is distributed*. In turn, group C will use two rules based on *the process for distributing the good*.

SET A: WHO BENEFITS?

- Divide the cake equally among class members.
- Divide the cake by the rank of class members. Define ranks according to the number of years of education of members.
- Divide the cake equally among two hypothetical groups: a group composed of 20 percent of the class members who have experienced historical discrimination in wages; and the remainder of the group. Each group must divide its equal portion in accordance with rank, as determined by years of education.

SET B: WHAT IS BEING DISTRIBUTED?

- Divide the cake so that persons with caloric intake below some standard receive larger pieces to compensate for their caloric deficit.

[49]Deborah Stone, *Policy Paradox: The Art of Political Decision Making*, rev. ed. (New York: W. W. Norton, 2002), Chapter 2: Equity, pp. 39–60, and Chapter 3: Efficiency, pp. 61–85.
[50]John Rawls, *A Theory of Justice* (Cambridge, MA: Harvard University Press, 1971).
[51]Robert Nozick, *Anarchy, State, and Utopia* (New York: Basic Books, 1974).

- Divide the cake so that persons with preferences for 3 pieces receive a larger piece than those with a preference for 2 pieces, who in turn receive a larger piece than those who prefer 1 piece or none.

SET C: WHAT IS THE PROCESS FOR DISTRIBUTING THE GOOD?

- Divide the cake according to a competitive process in which members in the first row of the room have a 100 percent chance of getting a piece of cake, members in the second row have an 80 percent chance, members in the third row have a 60 percent chance, members in the fourth row have a 40 percent chance, and members sitting further away have a 20 percent chance. Some members in the last row have no chance at all.
- Divide the cake according to a random process (lottery) in which each person has an equal probability of being selected to receive a piece of the cake. Repeat the process 30 times.[52]
- Divide the cake according to an electoral process in which a simple majority gets two-thirds of the cake. Choose two candidates and vote for them. ■

CASE 7.2 THE ECONOMICS OF MORALITY: EVALUATING LIVING WAGE POLICIES

Groups A, B, and C from the last exercise will apply their rules to the issue of whether or not the "living wage" policies described in Case 7.2 are justifiable. Justify each of your rules by drawing on normative ethical and meta-ethical theories described in this chapter, including Nozick and Rawls. Then construct a 10-minute briefing in which you present a recommendation(s) to the mayor of a city considering a living wage.

The living wage movement is considered by many to be one of the most important grass roots political mobilizations since the U.S. Civil Rights movement. Issues surrounding the adoption of a living wage are economic as well as moral and ethical. Economists are divided on the effects of the living wage on poverty, inequality, and employment, while the majority of religious and community leaders have supported the living wage for more than 100 years.

The following calculations have been extracted from a web page (www.livingwage.geog.psu.edu) developed by Dr. Amy K. Glasmeier of the Massachusetts Institute of Technology. The *living*

TABLE C7.2.1

Hourly Wages

Hourly Wages	One Adult	One Adult, One Child	Two Adults	Two Adults, One Child	Two Adults, Two Children
Living Wage	$11.86	$19.66	$16.29	$24.10	$30.30
Poverty Wage	$5.04	$6.68	$6.49	$7.81	$9.83
Minimum Wage	$7.25	$7.25	$7.25	$7.25	$7.25

[52]This is sampling with replacement, where each person selected is returned to the pool and can be selected again in each of the 30 draws.

wage shown is the hourly rate in 2011 that a full-time employed person (2,080 hours per year) in New York City (Queens County) must earn to support families of different sizes. In New York City, the *minimum wage* of $7.25 per hour is the same for all persons, regardless of the number of dependents. Although the poverty rate is typically quoted as gross annual income, for purposes of comparison it has been converted into an hourly wage.

Municipalities such as San Francisco, Detroit, and Chicago set the minimum wage higher than the federally mandated 2011 minimum wage of $7.25. The minimum living wage for an adult with one child and larger families also differs. In Pittsburgh, the living wage for an adult with one child is $34,244 per year, while in Los Angeles it is $45,235.

Typical Expenses

These figures were used to calculate the 2011 living wage in New York City. Numbers vary by family size. For typical expenses, living wages, and salaries for different occupations across the United States see www.livingwage.geog.psu.edu. ■

TABLE C7.2.2

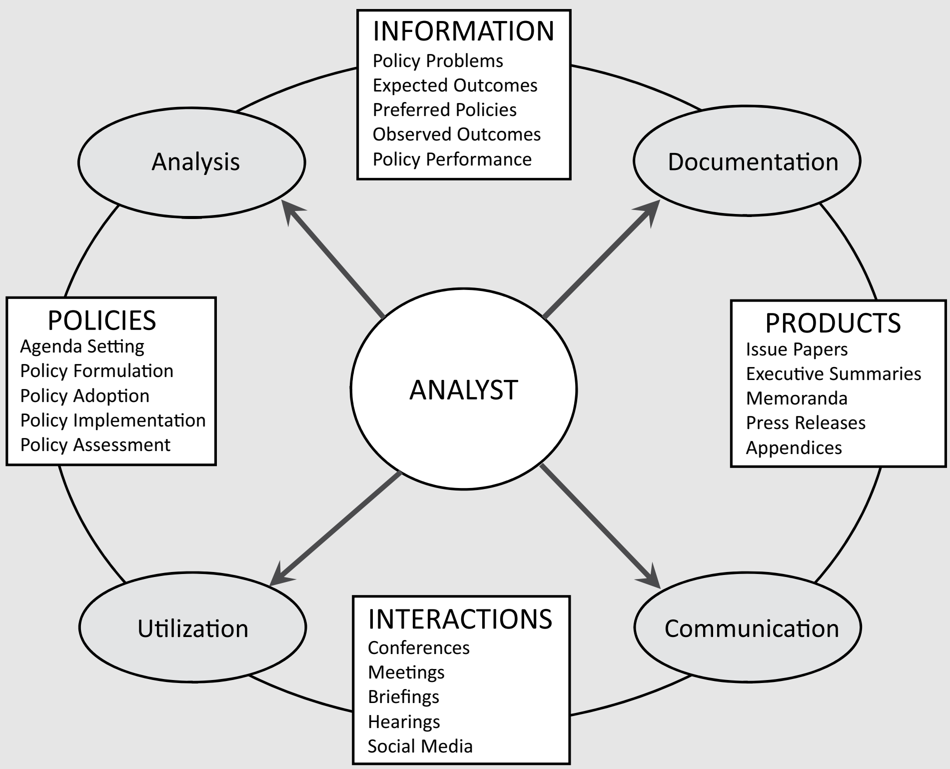
Monthly Expenses

Monthly Expenses	One Adult	One Adult, One Child	Two Adults	Two Adults, One Child	Two Adults, Two Children
Food	$232	$378	$448	$594	$740
Child Care	0	572	0	572	1,012
Medical	76	151	152	227	302
Housing	1,185	1,318	1,185	1,318	1,318
Transportation	232	397	464	629	794
Other	188	369	376	557	738
Monthly After-Tax Income That's Required	1,913	3,185	2,625	3,897	4,904
Annual After-Tax Income That's Required	22,956	38,220	31,500	46,764	58,853
Annual Taxes	1,706	2,678	2,387	3,373	4,161
Annual Before Tax Income That's Required	**$24,662**	**$40,898**	**$33,887**	**$50,137**	**$63,014**

Source: © 2011 Dr. Amy K. Glasmeier and The Pennsylvania State University.

PART III

Methods of Policy Communication



CHAPTER

8

Developing Policy Arguments

LEARNING OBJECTIVES

By studying this chapter you should be able to:

- Describe the structure of a policy argument
- Compare and contrast four types of policy claims
- Describe relations among seven elements of an argument
- Explain how objections and rebuttals affect the strength of an argument
- Demonstrate how qualifiers change in response to objections and rebuttals
- Distinguish and illustrate different modes of policy argumentation
- Explain why formal and informal fallacies diminish the strength of claims
- Use argument mapping techniques to represent cases of policy argumentation

INTRODUCTION

What policymakers understand and analysts often forget is that policy argumentation is central to policymaking.[1] Policy arguments are among the major vehicles for communicating policy-relevant information and an important source of knowledge about the ways policies are made and put into effect. Because policy arguments are carriers of policy-relevant information, they are also important for understanding the use and misuse of policy analysis by policymakers.[2]

The ability to organize, structure, and evaluate a policy argument is central to the process of critical analytical thinking introduced in the first part of this book. In contexts of practice, argumentation is not limited to the kinds of reasoning employed in formal logic or in the social sciences, for example, reasoning based on quantitative models designed to explain political behavior or on formal mathematical models of rational choice used in economics. Many other modes of argumentation—from ethical and political reasoning to reasoning by analogy and metaphor—coexist and compete for the attention of policymakers. Competent analysts should be able to compare and contrast different modes of reasoning and express technical arguments in a language that is comprehensible to policymakers.

THE STRUCTURE OF POLICY ARGUMENTS

A policy argument is the product of argumentation, which is the process. In real-life policy settings, arguments are complex and prone to misunderstanding. For this reason, conceptual models or maps are useful in identifying and relating the elements of policy arguments. One such model is the structural model of argument originated by the English-born philosopher Stephen Toulmin and extended by others.[3] Another structural model of argument called *Rationale*, developed by van Gelder and associates, is what we use here to map policy arguments (see Figure 8.1).[4]

Structural models are designed to investigate structures and processes of practical reasoning. Arguments based on practical reasoning, as distinguished from those based on the formal reasoning of mathematics and deductive logic, lack the certainty of formally valid logical arguments, for example, if A is preferred to B, and B is preferred to C, then A is preferred to C. Practical arguments are always uncertain, as are the reasons and evidence employed to justify conclusions. Reasons are not always stated explicitly, and in some cases,

[1]One of the landmark books on policy argumentation was Giandomenico Majone, *Evidence, Argument, and Persuasion in the Policy Process* (New Haven, CT: Yale University Press, 1992).
[2]This point is made emphatically in National Academy of Sciences, *Using Science as Evidence in Public Policy*. National Academies Press, 2012.
[3]Stephen Toulmin, *The Uses of Argument* (Cambridge: Cambridge University Press, 1958); and Stephen Toulmin, Robert Rieke, and Alan Janik, *An Introduction to Reasoning,* 2nd ed. (New York: Macmillan, 1984). Other models of reasoning and argument are Hayward Alker Jr., "The Dialectical Logic of Thucydides' Melian Dialogue," *American Political Science Review* 82, 3 (1988): 805–820; Michael Scriven, *Reasoning* (New York: McGraw Hill, 1977); D. R. Des Gasper, "Structures and Meanings: A Way to Introduce Argumentation Analysis in Policy Studies Education," *Africanus* (University of South Africa) 30, 1 (2000): 49–72; and D. R. Des Gasper, "Analyzing Policy Arguments," *European Journal of Development Research* 8, 1 (1996): 36–62.
[4]See Tim van Gelder, "The Rationale for Rationale." *Law, Probability, and Risk* 6 (2007): 23–42. The computer application called *Rationale* is available at www.austhink.com.

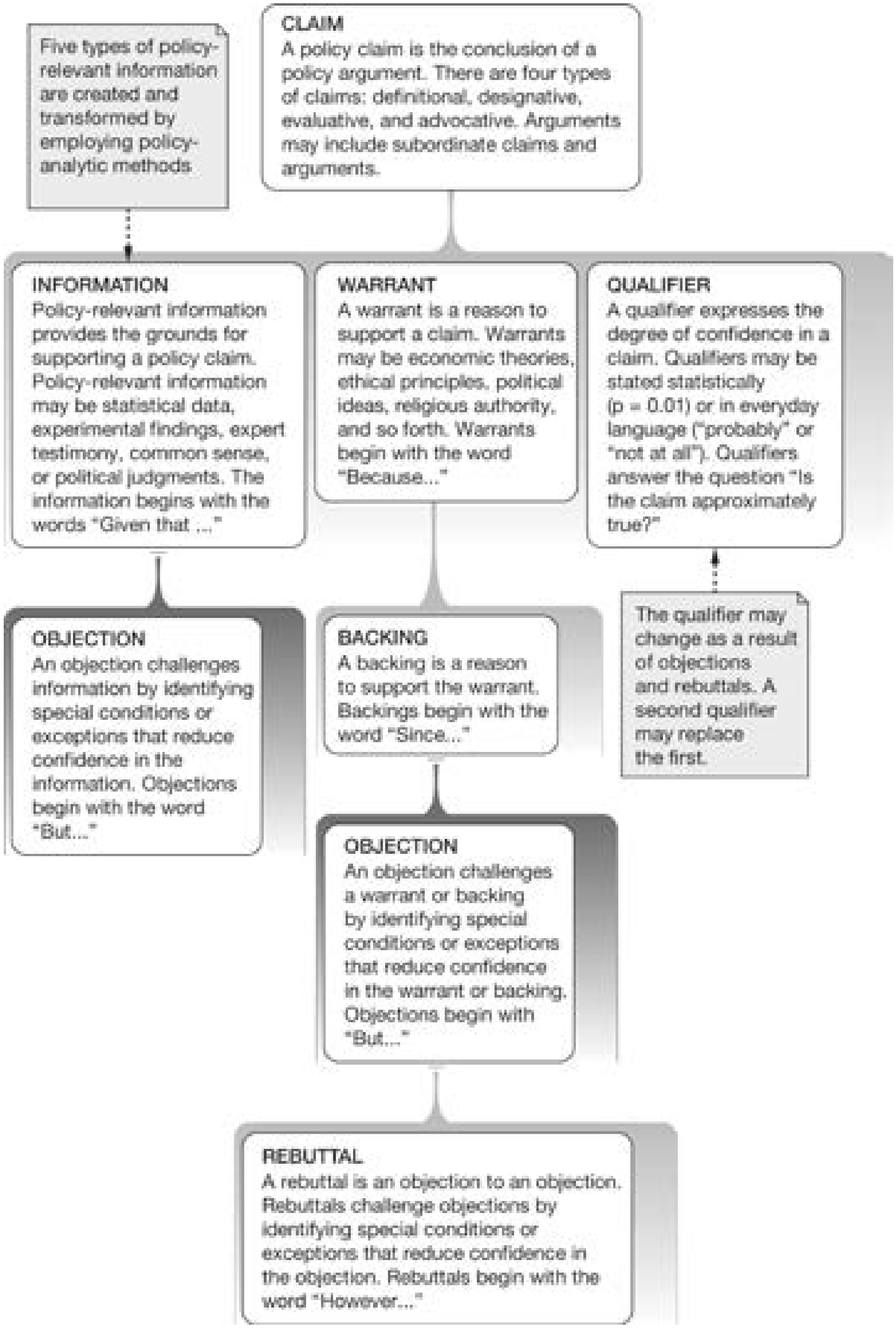


FIGURE 8.1
Structure of a Policy Argument

they are not stated at all. Even when they are stated, they are seldom complete or conclusive. In Toulmin's words, practical reasoning yields conclusions "about which we are not entirely confident by relating them back to other information about which we have greater assurance."[5] Policy arguments seldom if ever have the certainty of deductive logic.

Types of Knowledge Claims

A knowledge claim or contention is the conclusion of an argument. There are four types of knowledge claims: definitive, designative, evaluative, and advocative.

- *Definitive.* If a claim asserts that a policy has a particular definition, then the claim is definitive. Look for words such as *is*, *is not*, *constituted by*, *represented by*, *similar to*, and *different from*. A definitive claim states what something is or is not, what it is like or not like, or whether it belongs in one class or category rather than another. One type of definitive claim with particular relevance to policy is the metaphor. Policies have been defined metaphorically in terms of "war," "contagion," "epidemic," and "quarantine."
- *Designative.* If a claim asserts that some aspect of a policy has been or can be observed, or if observations are used to infer causes, then the claim is designative. Look for words that describe observed or observable characteristics such as *became*, *originated*, *linked*, *caused*, *effected*, *consequence*, and *prediction*. Some terms that seem to be designative are really evaluative, for example, "He is a good liberal" or "The evil empire (or axis of evil) is the enemy." Claims using these and other "appraising" descriptions are not questions of fact; they are evaluative, although it is also possible to study values empirically by making observations relevant to appraising descriptions. Good examples are measures of democracy, ethical behavior, conservatism, and liberalism included in questionnaires and interview schedules.[6] Designative claims rest on warrants that are "factual."
- *Evaluative.* If a claim asserts that some aspect of a policy has or does not have value or worth, it is evaluative. Look for words such as *good*, *bad*, *right*, *wrong*, *beneficial*, *costly*, *efficient*, *responsive*, *equitable*, *just*, *fair*, and *secure*. Examples include "The policy will bring about a 'living wage,' thus moving society toward equity and responsiveness," "The program is less efficient than expected," or "The policy will build a healthy economy." Some evaluative claims refer to states or conditions (e.g., a just society) and some to procedural processes (e.g., a fair trial). Evaluative claims rest on warrants that involve values, ethics, and meta-ethics.
- *Advocative.* If a claim asserts that a government or division within it should take action, the claim is advocative. Look for words like *should*, *needs to*, and *must*. Examples include "The World Bank should terminate its structural adjustment program," "Congress should pass the Equal Rights Amendment," "The United States should sign the Kyoto Treaty," or "Auto manufacturers should produce more fuel

[5]Toulmin, *The Uses of Argument*, p. 127.
[6]See Delbert C. Miller and Neil J. Salkind, *Handbook of Research Design and Social Measurement*, 6th ed. (Newbury Park, CA: Sage Publications, 2002).

efficient vehicles." Advocative claims rest on warrants that simultaneously involve facts and values.

Policy Maps

Policy maps are useful in representing complex arguments. Arguments have seven elements: claim, *C*, information, *I,* warrant, *W,* backing, *B,* objection, *O*, rebuttal, *R,* and qualifier, *Q* (Figure 8.1). The claim, *C,* is the conclusion or contention of an argument. The claim, *C,* is supported by policy-relevant information, *I,* which is the beginning of an argument. The warrant, *W,* is the reason for concluding *C* from *I.* The qualifier, *Q*, expresses the approximate truth, plausibility, or degree of confidence in the claim, *C.* Consider the following example:

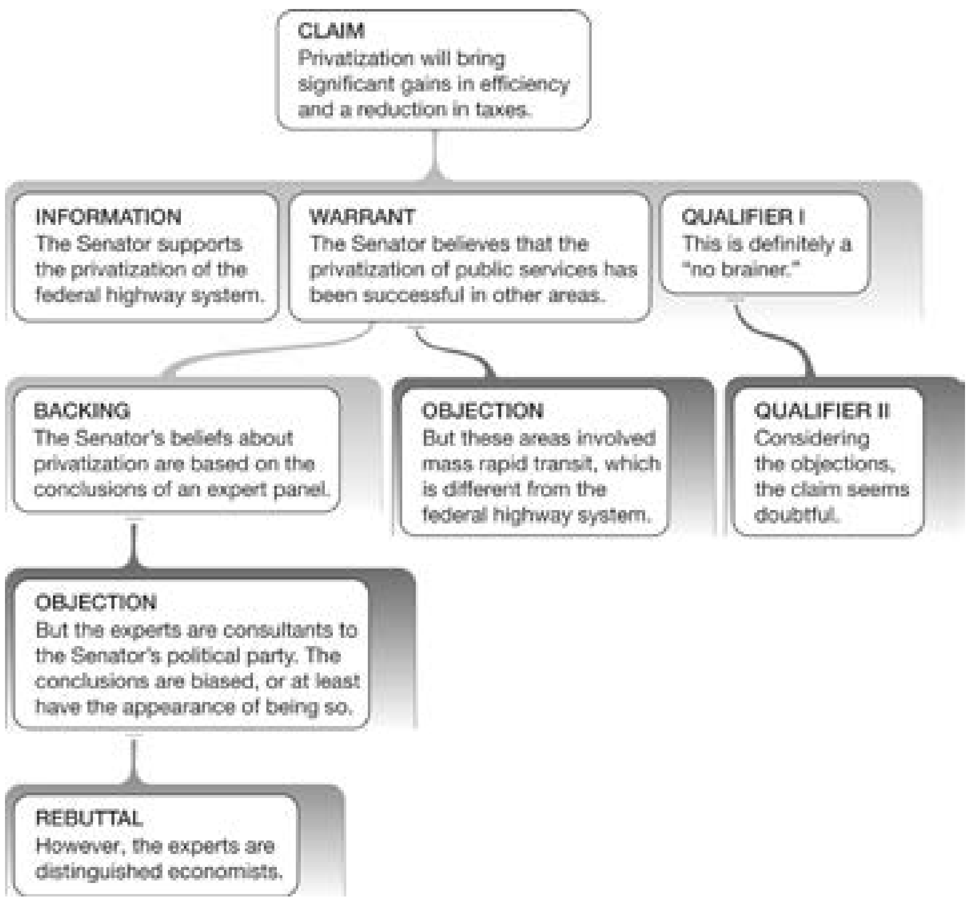


FIGURE 8.2
Argument Map—Privatizing Transportation

> The Senator supports the privatization of the federal highway system, which will bring gains in efficiency and reduced taxes. Considering that the privatization of public services has been successful in other areas, this is definitely a "no brainer." This same conclusion was reached by a panel of experts on privatization.

The Senator's argument may be broken down into its basic elements (Figure 8.2):

> The Senator supports the privatization of the federal highway system (INFORMATION), which will bring significant gains in efficiency and a reduction in taxes (CLAIM). Considering the WARRANT that the privatization of public services has been successful in other areas, QUALIFIER I states that "this is definitely a 'no brainer.' "

Additional elements are then introduced to strengthen the argument. For instance, a BACKING has been added ("This is the conclusion of an expert panel on privatization."). Finally, an OBJECTION to the backing states that the experts are employed by the Senator's party. A weak REBUTTAL then states that the experts are distinguished economists. The rebuttal is weak because it is not a plausible challenge to the objection, which is about politics and not academic standing. A second OBJECTION states that the management of urban mass transit is different than managing the federal highway system. This appears to be a plausible challenge. Finally, QUALIFIER I was that the claim is a "no brainer." In QUALIFIER II, this is reduced to "perhaps," considering the strength of the objections and rebuttal. This considerably weakens the CLAIM that "the privatization of the federal highway system . . . will bring significant gains in efficiency and a reduction in taxes."

EXHIBIT 8.1

Identifying and Arranging Elements of a Policy Argument

Policy arguments have seven elements: information, claim, qualifier, warrant, backing, objection, and rebuttal. The following guidelines are useful in identifying and arranging these elements:

1. If possible, do a stakeholder analysis (see Chapter 3) before you analyze an argument. This is the main source of elements.
2. Start by locating the claim, which is the endpoint or output of the argument. A claim is always more general than the information on which it is partly based. A claim usually involves an "inferential leap" beyond information.
3. Look for language that indicates the degree of credibility the arguer attaches to the claim—this is the qualifier.
4. Look for the information that supports the claim. The information answers two questions: "What does the arguer have to go on?" and "Is it relevant to the case at hand?"
5. Look for the warrant, which in conjunction with the information supports the claim. The warrant answers the question: "Why is the arguer justified in making the claim on the basis of the information?"
6. Repeat the same procedure with the backing. If there is a question whether a statement is a backing or a warrant, look for the one that is more general. This is the backing.
7. Remember that a warrant or backing may be implicit and unstated—do not expect arguments to be entirely transparent.

(*continued*)

EXHIBIT 8.1 (Continued)

8. Look to the arguments of other stakeholders to find objections and rebuttals. Objections and rebuttals require someone who actually believes in them.
9. Remember that elements can be rules, principles, statements, or entire arguments.
10. An isolated argument is static; argumentation, which involves at least two contending stakeholders, is dynamic. The initial qualifier usually changes when objections and rebuttals challenge the argument.
11. Most qualifiers become weaker, although some stay the same. Some can grow stronger (*a fortiori*) by withstanding challenges.
12. Argumentation produces "trees" and "chains" involving processes of argumentation over time among many stakeholders.

MODES OF POLICY ARGUMENTATION

Distinct modes of argumentation are used to justify policy claims. Modes of argumentation, which are specific patterns of reasoning about policy, include reasoning from authority, method, generalization, classification, intuition, cause, sign, motivation, analogy, parallel case, and ethics.[7] Each of these modes of argumentation and its characteristic reasoning pattern is described in Table 8.1. Note that more than one mode may be used in a complex policy argument.

Argumentation from Authority

In the authoritative mode, claims are based on arguments from authority. Policy-relevant information consists of factual reports or expressions of opinion. The function of the warrant is to affirm the reliability or trustworthiness of the source of the information. Depending on the social context, authorities may be kings, magicians, or religious leaders, or they may occupy roles as presidents, legislators, scientists, professors, or news reporters.

In an authoritative argument the claim reiterates the policy-relevant information that has been provided by the authority, whose reliability, status, or sagacity has been underwritten by the warrant. To illustrate (Figure 8.3), let us imagine that a policy analyst advising the National Security Council at the height of the 1999 U.S.–NATO attack on Serbia made the designative claim, *C*: "although Western leaders continue to deny it, there can be little doubt that the bombing campaign has provided both motive and opportunity for a wider and more savage Serbian operation than what was first envisioned."[8] The information, *I*, is from a statement by Carnes Lord, a professor at the Fletcher School of Law and Diplomacy

[7]Some of the modes presented here draw from Wayne Brockriede and Douglas Ehninger, "Toulmin or Argument: An Interpretation and Application," *Quarterly Journal of Speech* 1006 (1960): 45–53, and Toulmin, Rieke, and Janik, *An Introduction to Reasoning*, 2nd ed., pp. 213–237. I have added several additional modes.
[8]*Boston Globe*, April 4, 1999. Quoted in Noam Chomsky, *Rogue States: The Rule of Force in World Affairs* (Cambridge, MA: South End Press, 2000), p. 35.

TABLE 8.1

Modes of Policy Argumentation with Reasoning Patterns

Mode	Reasoning Pattern
Authority	Reasoning from authority is based on warrants having to do with the achieved or ascribed statuses of producers of policy-relevant information, for example, experts, insiders, scientists, specialists, gurus, power brokers. Footnotes and references are disguised authoritative arguments.
Method	Reasoning from method is based on warrants about the approved status of methods or techniques used to produce information. The focus is on the achieved or ascribed status or "power" of procedures, rather than persons. Examples include approved statistical, econometric, qualitative, ethnographic, and hermeneutic methods and techniques.
Generalization	Reasoning from generalization is based on similarities between samples and populations from which samples are selected. Although samples can be random, generalizations can also be based on qualitative comparisons. The assumption is that what is true of members of a sample will also be true of members of the population not included in the sample. For example, random samples of $n \geq 30$ are taken to be representative of the (unobserved and often unobservable) population of elements from which the sample is drawn.
Classification	Reasoning from classification has to do with membership in a defined class. The reasoning is that what is true of the class of persons or events described in the warrant is also true of individuals or groups which are members of the class described in the information. An example is the untenable ideological argument that because a country has a socialist economy it must be undemocratic, because all socialist systems are undemocratic.
Cause	Reasoning from cause is about the activity of generative powers ("causes") and their consequences ("effects"). For example, a claim may be made based on general propositions, or laws, of economics that state invariant relations between cause and effect. Other causal claims are based on observing conditions that must be satisfied to infer that a policy has a specified effect. Most argumentation in the social and natural sciences is based on reasoning from cause.
Sign	Reasoning from sign is based on signs, or indicators, and their referents. The presence of a sign indicates the presence of an event or condition, because the sign and what it refers to occur together. Examples are indicators of institutional performance such as "organizational report cards" and "benchmarks" or indicators of economic performance such as "leading economic indicators." Signs are not causes, because causality must satisfy additional requirements not expected of signs.
Motivation	Reasoning from motivation is based on the motivating power of goals, values, and intentions in shaping individual and collective behavior. For example, a claim that citizens will support the strict enforcement of pollution standards might be based on reasoning that, since citizens are motivated by the desire to achieve the goal of clean air and water, they will act to offer their support.

(*continued*)

Mode	Reasoning Pattern
Intuition	Reasoning from intuition is based on the conscious or preconscious cognitive, emotional, or spiritual states of producers of policy-relevant information. For example, the awareness that an adviser has some special insight, feeling, or "tacit knowledge" may serve as a reason to accept his judgment.
Analogy	Reasoning from analogies is based on similarities between relations found in a given case and relations characteristic of a metaphor or analogy. For example, the claim that government should "quarantine" a country by interdicting illegal drugs—with the illegal drugs seen as an "infectious disease"—is based on reasoning that, since quarantine has been effective in cases of infectious diseases, interdiction will be effective in the case of illegal drugs.
Parallel Case	Reasoning from parallel case is based on similarities among two or more cases of policymaking. For example, a reason that a local government should strictly enforce pollution standards is that a parallel policy was successfully implemented in a similar local government elsewhere.
Ethics	Reasoning from ethics is based on judgments about the rightness or wrongness, goodness or badness, of policies or their consequences. For example, policy claims are frequently based on moral principles stating the conditions of a "just" or "good" society, or on ethical norms prohibiting lying in public life. Moral principles and ethical norms go beyond the values and norms of particular individuals or groups. In public policy, many arguments about economic benefits and costs involve unstated or implicit moral and ethical reasoning.

and former Bush administration national security advisor. The warrant *W* affirms Lord's reliability and is backed *B* by an additional authoritative argument intended to add to the persuasiveness of the argument. The objection *O* challenges the initial warrant, but without weakening the claim's credibility, which is stated in the qualifier *Q*. The rebuttal *R* has virtually no effect on *Q*, which does not change in response to the challenge by a rival authority, State Department spokesman James Rubin.

Argumentation from Method

Argumentation from method is based on warrants about the approved status of methods or techniques used to produce information. Policy-relevant information may consist of factual statements or reports. The role of the warrant is to provide a reason for accepting the claim based on the information by associating the latter with the use of an approved method or rule. Usually, the claim is that the event, condition, or object described in the information should be regarded as valuable (or worthless), because of the method used to produce it. Consider the following public investment problem. An analyst has information *I* that the production of energy per dollar is greater in nuclear power plants than in hydroelectric plants, which in turn produce more energy per dollar than solar power plants. The claim *C*

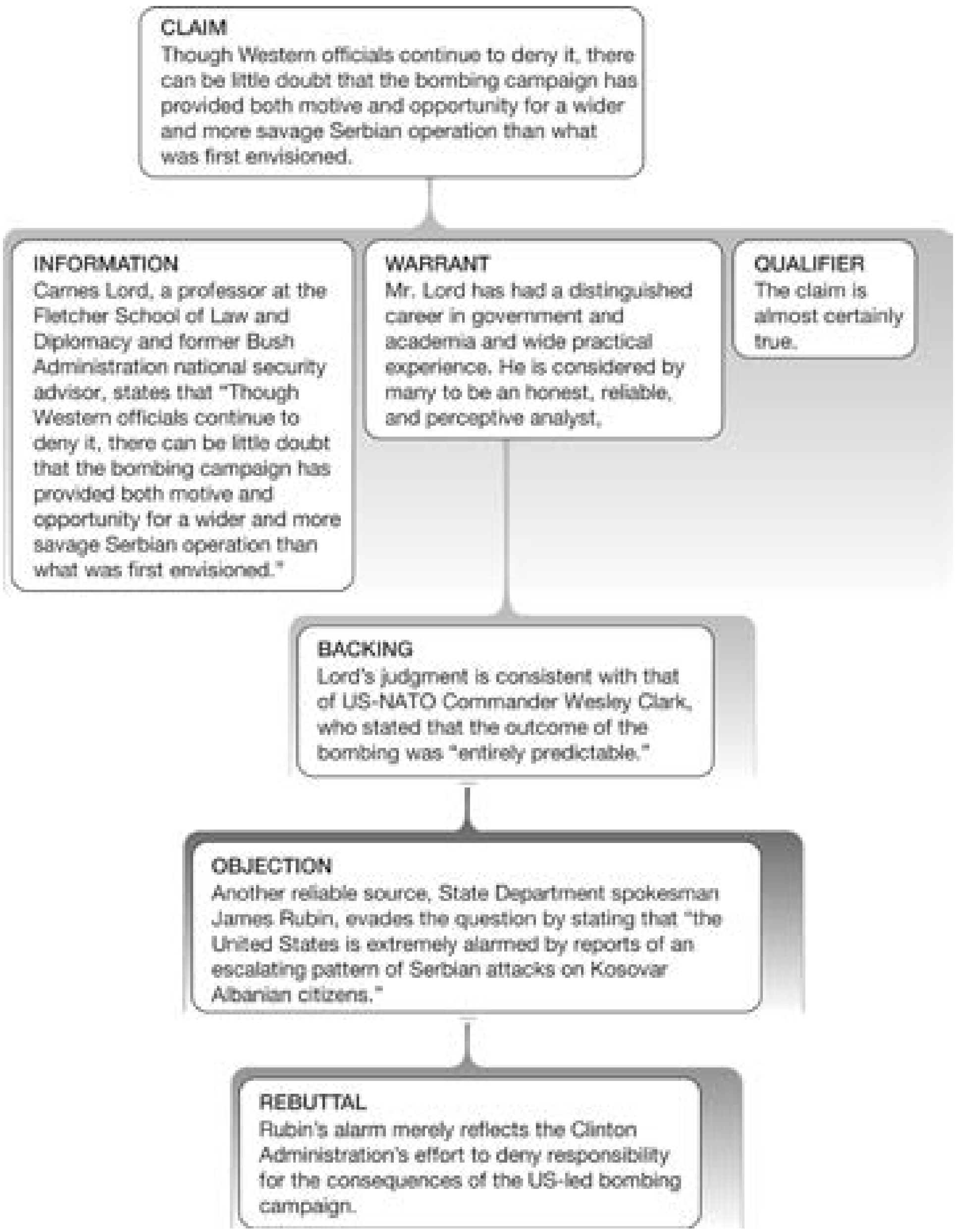


FIGURE 8.3
Argumentation from Authority—Unintended Consequences of the U.S.–NATO Attack on Serbia

is that the government should invest in nuclear energy rather than solar energy. The warrant W associates the information I with claim C by invoking the transitivity rule of mathematical economics.[9] The warrant is backed B by the presumption that transitivity is a "universal selection rule" that guarantees the rationality of choice.

The rebuttal R challenges the presumption about the universal validity of transitivity by pointing to the presence of cyclical preferences. The original qualifier $Q1$ is reduced from "very likely" to the new qualifier $Q2$ which is now stated as "quite uncertain" (Figure 8.4).

In argumentation from method, claims are assessed in terms of the achieved or ascribed status of general procedures (methods) and specialized ones (techniques), along with the rules guiding their use. The methods and techniques can be "analytic," in the sense of the dictionary definition of analysis as the separation or breaking up of a whole into its

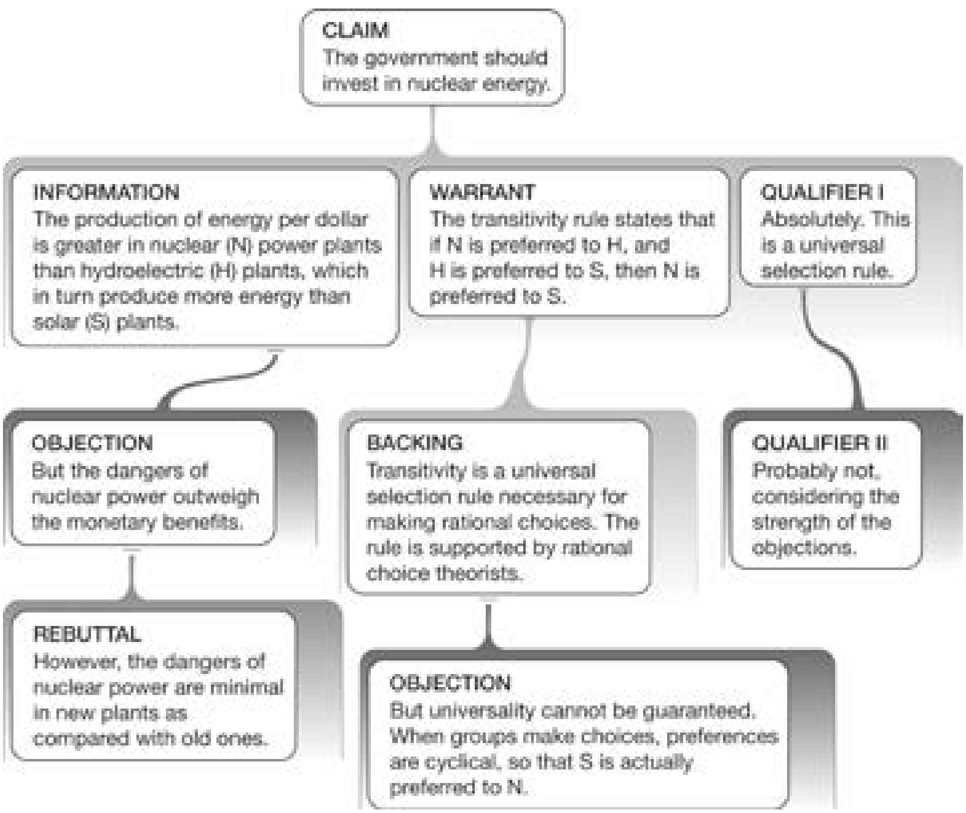


FIGURE 8.4
Argumentation from Method—Intransitivity of Preferences for Nuclear Power

[9] On rules expressing transitive and cyclical preferences, see, for example, Norman Frohlich and Joe A. Oppenheimer, *Modern Political Economy* (Englewood Cliffs, NJ: Prentice Hall, 1978), pp. 6–13.

fundamental elements or constituent parts. Policy analysts who accept the authority of analytic methods such as econometrics, benefit–cost analysis, or decision analysis sometimes seem to believe that the use of such methods actually "sets the policy agenda and its directions, that useful analysis will be used analysis."[10]

The authority of methods need not be derived from rules of formal logic or mathematics, as the history of qualitative methods shows. Many qualitative methods originate in the hermeneutic tradition, which evolved from the interpretation of biblical texts. The authority of qualitative methods, like any other, stems from the professional and scientific communities that create definitions of the purpose, scope, and proper application of approved methods.[11] These communities are extra-scientific sources of authority.

The transitivity rule and other basic axioms of mathematics have also arisen from extra-scientific sources:

> Consider the axiom which asserts the transitivity of preference: if A is preferred to B, and B to C, then A is (or rationally must be) preferred to C. The intuitive appeal of this assertion is so great that few if any economists feel the urge to build formal economic systems in which the axiom fails. Geometry . . . is the classic example; the intuitive strength of Euclid's "postulates" was so great that for two thousand years geometers played their games strictly within the domain . . . which Euclid had laid down. Even when non-Euclidean geometries were discovered in the early nineteenth century, most mathematicians never thought of them as valid.[12]

Adherence to approved methods is wrongly believed to make policy decisions somehow more "rational." A rational choice is thought to be possible if the analyst can order all consequences associated with action, if the ordering of consequences is transitive, and if the analyst can consistently and in a transitive fashion choose the alternative that will bring the greatest benefit in relation to cost.[13] Challenges to this argument may be made on empirical, authoritative, intuitive, and ethical grounds.[14] New schools of analysis may serve as a source of approved methods, new axioms providing for nontransitive preferences may come to be accepted on intuitive grounds and rules that run counter to moral principles and ethical norms may be replaced with new ones.[15] Challenges may also be made on pragmatic grounds, for example, by arguing that the transitivity rule does not actually promote better decisions in policy settings characterized by incomplete information, value conflicts, multiple competing objectives, partisan mutual adjustment, and "organized anarchy."[16] In

[10]Allen Schick, "Beyond Analysis," *Public Administration Review* 37, 3 (1977): 259.

[11]Approved methods are part of Kuhn's disciplinary matrix. Thomas Kuhn, *The Structure of Scientific Revolutions,* 2nd ed. (Chicago, IL: University of Chicago Press, 1971), p. 103.

[12]C. West Churchman, *The Design of Inquiring Systems: Basic Concepts of Systems and Organization* (New York: Basic Books, 1971), p. 25.

[13]Joseph L. Bower, "Descriptive Decision Theory from the 'Administrative' Viewpoint," in *The Study of Policy Formation,* ed. Raymond A. Bauer and Kenneth J. Gergen (New York: Free Press, 1968), pp. 104–106.

[14]Good examples of these challenges may be found in Jeffrey Friedman, ed., *The Rational Choice Controversy: Economic Models of Politics Reconsidered* (New Haven, CT: Yale University Press, 1996).

[15]See the discussion of the evolution of welfare economics in public policy analysis by Duncan MacRae Jr., *The Social Function of Social Science* (New Haven, CT: Yale University Press, 1976), pp. 107–157.

[16]See Chapter 2.

the last analysis, a successful argument from method is a pragmatic matter. An argument from method must demonstrate only that the results of using particular rules are superior to those that occur without them and that the observed improvement is a consequence of using the rule or procedure.[17]

Argumentation from Generalization

Arguments from generalization involve samples. Policy-relevant information consists of sampled elements—events, behaviors, persons, groups, organizations—which are taken to be representative of a larger population of the same elements. The function of the warrant is to affirm that what is true of the sampled elements is also true of the unobserved (and often unobservable) elements in the population. The policy claim rests on the argument that the sample is a satisfactory representation of the population.

To illustrate, consider the director of a community food bank who wants to know whether persons receiving food are getting an adequate daily allowance of calcium, one of the most important minerals in the body (Figure 8.5). The director, who is attentive to reasons that might justify additional funding for the food bank, makes the claim *C* that it is "pretty likely" that food bank clients are receiving sufficient calcium—specifically, given that the observed amount of 755 milligrams (mg) of calcium could have occurred by chance 9 times out of 100, it is not reliably different than the Recommended Daily Allowance (RDA) of 800 mg, an amount prescribed by the Food and Nutrition Board of the National Academy of Sciences.

In this case, the information *I* describes the average daily intake of calcium (755 mg) measured in a random sample (sometimes inappropriately called a "scientific sample" in policy circles) of fifty clients. The information *I* indicates that the difference between 755 mg and 800 mg could occur by chance 9 times out of 100, a conclusion reached on the basis of a statistical test (a "two-sample test"). The probability value ($p = 0.09$) is included in the qualifier *Q* in the ordinary language of "pretty likely." The warrant *W* that provides the reason for moving from information *I* to claim *C* has two parts: the rule that a random sample of at least 30 is adequate in such cases to generalize to the population from which the sample is selected and the practice of accepting a level of statistical significance of $p = 0.05$ (5 times out of 100) as an "acceptable" level of risk. When pressed for additional justification, the director checks her statistics text to find the appropriate theoretical backing *B* for the rule $n \geq 30$. The backing *B* is the Central Limit Theorem of probability theory.

A member of the director's staff responsible for distributing the food is sensitive to criticism and resistant to the idea that clients served by the food bank may have a calcium deficiency. He therefore challenges the claim with several rebuttals *R*: A 9 percent margin of error leaves too much room for chance, including a 91 percent (100 – 9) chance of a "false negative" conclusion of no difference (Type II error); another random sample could result in another conclusion; and, in any case, the difference between 755 and 800 mg of calcium is *practically* significant and should be looked at carefully. The rebuttals are very plausible,

[17]Bower, "Descriptive Decision Theory from the 'Administrative' Viewpoint," p. 106. See Nicholas Rescher's essays on methodological pragmatism, for example, *Induction* (Pittsburgh, PA: University of Pittsburgh Press, 1980).

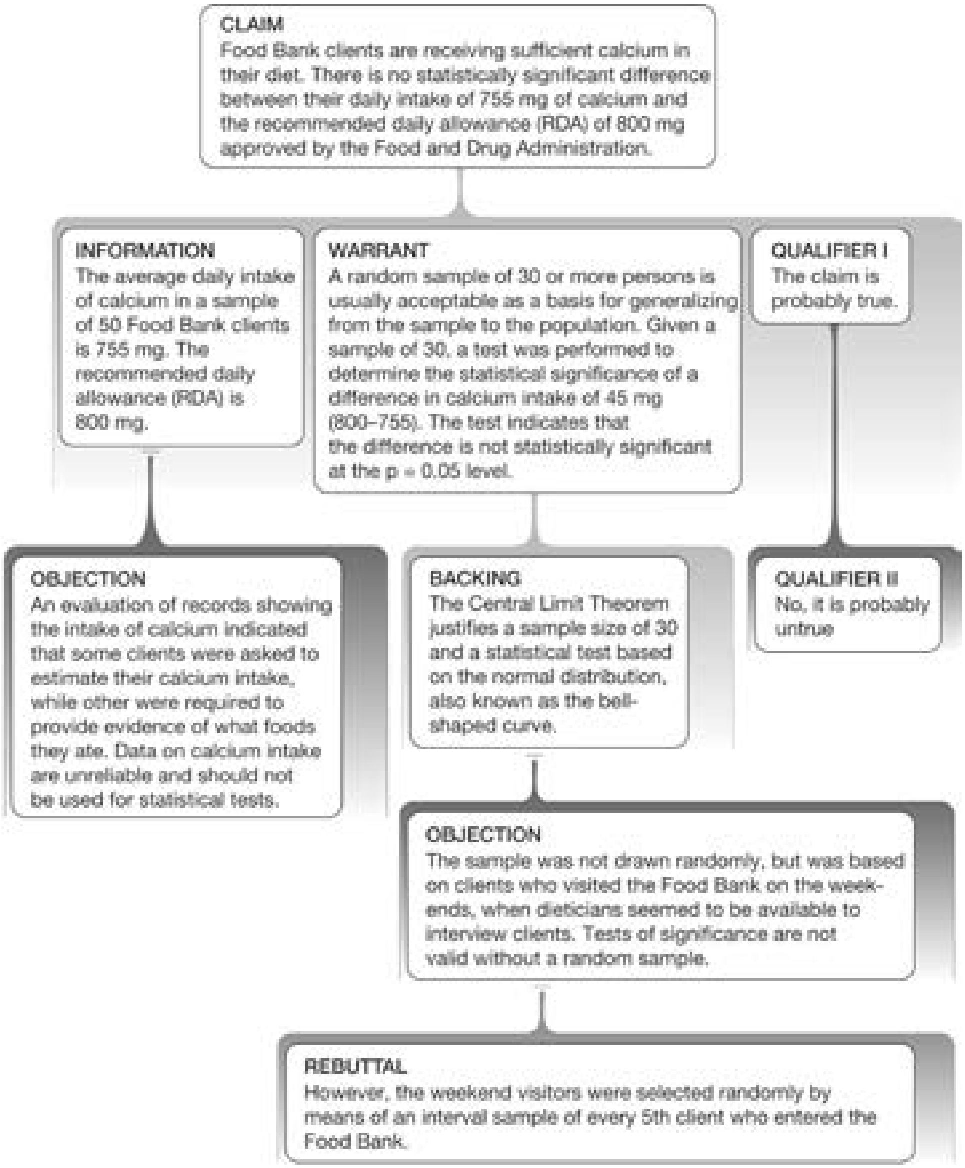


FIGURE 8.5
Argumentation from Generalization—A "False Positive" About the Success of Community Nutrition

and we would guess that the director's qualifier *Q1* changed from "pretty likely" to *Q2* "not likely" that clients are receiving their minimum RDA of calcium. On balance, the director's original claim—that calcium intake is probably adequate—is diminished by the rebuttals of the well-meaning and critical staff member.

Arguments from generalization are not always statistical, in the specific sense that statistics are estimates of population values (called parameters). Nonrandom samples—for example, purposive samples, theoretical samples, and sociometric (snowball) samples—do not permit statistical estimation. They are nevertheless useful in making claims about populations.[18] Even case studies (where $n = 1$) may be used to generalize to wider populations by means of various pattern-matching methods.[19]

Argumentation from Classification

Argumentation from classification focuses on membership in a defined class. The reasoning is that what is true of the class of persons or events described in the warrant is also true of individuals or groups which are members of the class, as they are described in the information. To illustrate, consider the following argument about the relationship between regime type and the control of terrorism (Figure 8.6). The information *I* is Iran, Iraq, and the Palestinian Authority are authoritarian dictatorships. The claim *C* is that Iran, Iraq, and the Palestinian Authority control terrorism within their borders. The warrant *W* is that authoritarian regimes exercise firm control of terrorists and other armed groups within their territories. The backing *B* has two parts. The first is that firm control exercised by authoritarian dictatorships is possible because such regimes do not permit the private use of weapons without the express or tacit approval of the government. The second is that the theory of authoritarian dictatorship was developed and tested by distinguished scholars at Harvard University. The objections *O* are that one or more of the Middle Eastern regimes permit the private use of weapons and that the definition of "authoritarian regime" may be inappropriate. The original qualifier *Q1* (certainly) is reduced to *Q2* (it is not clear).

The plausibility of classificational arguments depends on the completeness and internal consistency of the properties employed to define a class. Various classes of political regimes—authoritarian dictatorships, totalitarian democracies, socialist democracies, capitalist democracies—are less homogeneous and internally consistent than popular classifications suggest. The same is true for classes of policies (e.g., "privatization"); organizations (e.g., "bureaucracy"); political doctrines (e.g., "liberal" and "conservative"); and groups (e.g., "lower class," "middle class," "upper class"). Many apparently simple classifications turn out to be complex, not simple, and they may be ideologies in disguise.

---

[18]See, for example, Delbert C. Miller, *Handbook of Research Design and Social Measurement*, 4th ed. (Newbury Park, CA: Sage Publications, 1991). Generalizability (external validity) in experimental research is essentially nonstatistical. Among other reasons, this is why Campbell called it "proximal similarity." See William R. Shadish, Thomas D. Cook, and Donald T. Campbell, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference* (Belmont, CA: Wadsworth, Cengage Learning, 2002).

[19]See William N. Dunn, "Pattern Matching: Methodology," *International Encyclopedia of the Social and Behavioral Sciences* (New York: Elsevier, 2002).

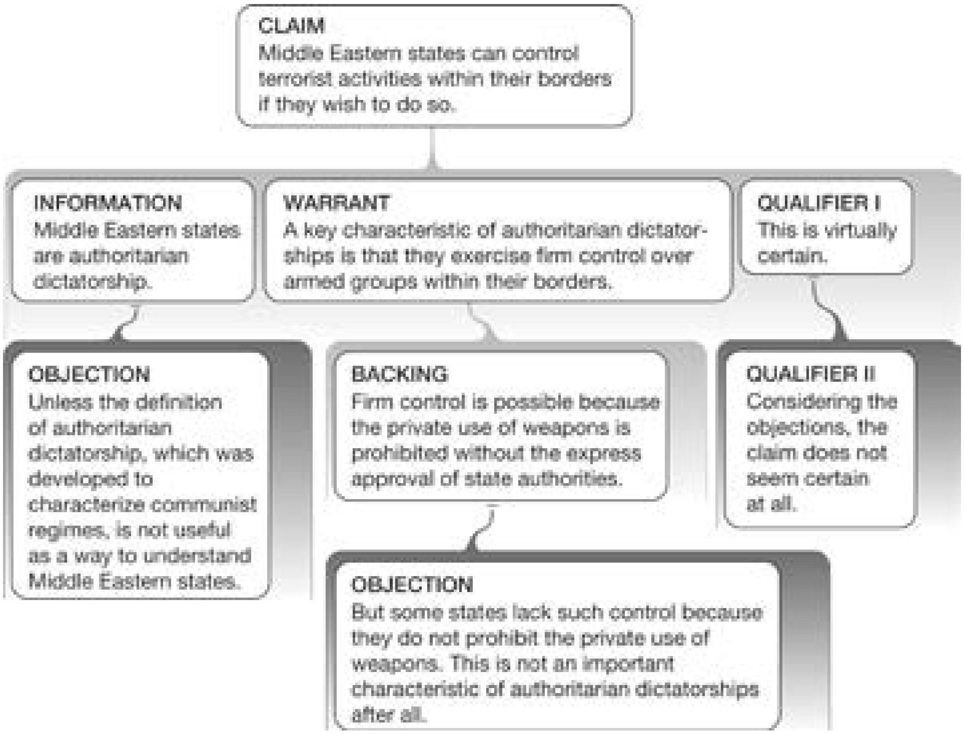


FIGURE 8.6
Argumentation from Classification—Challenging Claims about Authoritarian Rule and Terrorism

Argumentation from Cause

Argumentation from cause focuses on the causes and effects of public policies.[20] In causal arguments, information consists of one or more evidently factual statements or reports about a policy environment, a policy stakeholder, or a policy. The warrant transforms

[20]Causal argumentation dominates efforts of political scientists to explain policymaking. Recent examples include contributors to Sabatier, *Theories of the Policy Process*. Other examples are James E. Anderson, *Public Policy-Making* (New York: Praeger Publishers, 1975); Thomas R. Dye, *Understanding Public Policy,* 3rd ed. (Englewood Cliffs, NJ: Prentice Hall, 1978); Robert Eyestone, *The Threads of Public Policy: A Study in Policy Leadership* (Indianapolis, IN: Bobbs-Merrill, 1971); Jerald Hage and J. Rogers Hollingsworth, "The First Steps toward the Integration of Social Theory and Public Policy," *Annals of the American Academy of Political and Social Science,* 434 (November 1977): 1–23; Richard I. Hofferbert, *The Study of Public Policy* (Indianapolis, IN: Bobbs-Merrill, 1974); Charles O. Jones, An *Introduction to the Study of Public Policy,* 2nd ed. (North Scituate, MA: Duxbury Press, 1977); Robert L. Lineberry, *American Public Policy: What Government Does and What Difference It Makes* (New York: Harper & Row, 1977); Austin Ranney, ed., *Political Science and Public Policy* (Chicago, IL: Markham, 1968); Richard Rose, ed., *The Dynamics of Public Policy: A Comparative Analysis* (Beverly Hills, CA: Sage Publications, 1976); Ira Sharkansky, ed., *Policy Analysis in Political Science* (Cambridge, MA: Markham, 1970); and Peter Woll, *Public Policy* (Cambridge, MA: Winthrop Publishers, 1974).

these statements or reports by relating them to generative powers (causes) and their results (effects). The claim relates these causes and effects back to the information supplied.

The role of causal arguments in transforming policy-relevant information into policy claims may be illustrated by Allison's well-known causal explanations of foreign policy behavior during the Cuban missile crisis in October 1962.[21] Showing how different models yield alternative explanations of foreign policy, Allison argues that (1) government policy analysts think about problems of foreign policy in terms of implicit conceptual models that shape their thought; (2) most analysts explain the behavior of governments in terms of one basic model, one that assumes the rationality of political choices (*rational policy model*); and (3) alternative models, including those that emphasize organizational processes (*organizational process model*) and bureaucratic politics (*bureaucratic politics model*), provide bases for improved explanations of foreign policy behavior.

In contrasting alternative models, Allison assesses U.S. foreign policy during the Cuban missile crisis by reviewing explanatory arguments derived from the three conceptual models. In 1962, the policy alternatives open to the United States ranged from no action and diplomatic pressures to secret negotiations, invasion, surgical air strike, and blockade. The alternatives are examined in terms of rival explanations of foreign policy behavior. The structure of these explanations conforms to a type of causal explanation that philosophers call deductive-nomological (D-N) explanation, which holds that valid explanations are possible only when general theoretical propositions or laws link prior circumstances with subsequent events.[22]

Among the several advocative claims made at the time of the Cuban missile crisis, let us consider the policy recommendation actually adopted by the United States: "The United States should blockade Cuba." In this case, the policy-relevant information *I* is "The Soviet Union is placing offensive missiles in Cuba." To carry information *I* to claim *C,* a warrant *W* answers the question: Why is it plausible, given the information provided, to claim that the United States should blockade Cuba? The warrant provides the answer by stating *since* "the blockade will force the withdrawal of missiles by showing the Russians that the United States is determined to use force." In providing additional reasons to accept the claim, the backing *B* answers the question: Why would the blockade have this effect? The backing answers the question by stating *because* "[a]n increase in the cost of an alternative reduces the likelihood of that alternative being chosen."[23] The backing *B* represents a general theoretical proposition, or law, within the rational policy model (Figure 8.7). In this case, *Q1* (probably) changes to *Q2* (probably not) after the objection *O* has successfully challenged the backing *B* of the warrant *W.*

The primary purpose of Allison's account is not to demonstrate the inherent superiority of one or another of the three explanatory models. It is rather to show that the use of multiple competing models can result in improved explanations of foreign policy behavior. The use of multiple models moves policy analysis from a self-contained and static single argument about the relation between information and claim to a new stage of dynamic debate. In this context, the organizational, process model provides an objection *O* with

[21]Graham T. Allison, "Conceptual Models and the Cuban Missile Crisis," *American Political Science Review* 3002, 3 (1969): 689–718.

[22]See Carl G. Hempel, *Aspects of Scientific Explanation* (New York: Free Press, 1965).

[23]Allison, "Conceptual Models," p. 694.

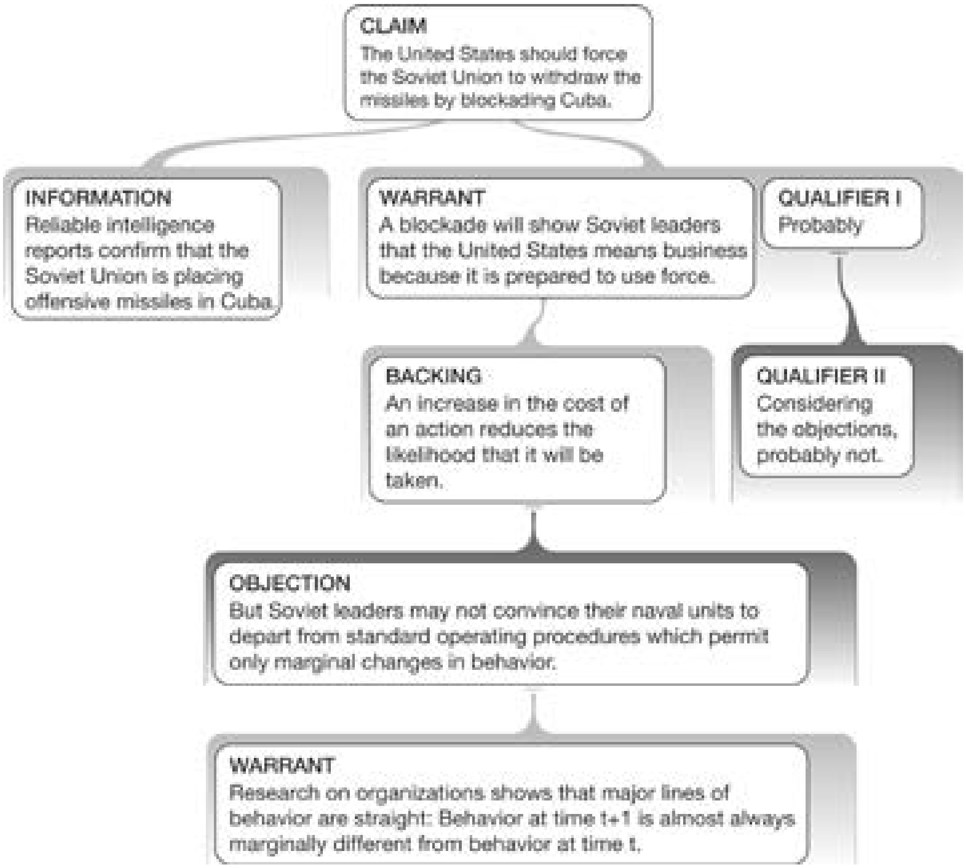


FIGURE 8.7
Argumentation from Theoretical Cause—Competing Deductive-Nomological Explanations of the Cuban Missile Crisis

its own backing *B*, in the form of a competing causal argument. The objection *O* states that Soviet leaders may be unable to force their own organizational units to depart from assigned tasks and routines. This can be expected to occur *because* "[m]ajor lines of organizational behavior are straight, that is, behavior at one time is marginally different from that same behavior at $t - 1$."[24] The backing *B* for the objection *O* is again a general proposition or law within the organizational process model.

The case of the Cuban missile crisis illustrates some of the limitations of causal argumentation based on D-N explanation. First, several competing causal arguments are equally compatible as general propositions or laws. These arguments, each backed by scientific theories, cannot be confirmed or refuted solely on the basis of this or any other information

[24]Ibid., p. 702.

or data.[25] Second, a given causal argument, however persuasive, cannot directly lead to an advocative claim or recommendation, because traditional causal explanations do not themselves contain value premises.[26] In the following example, there is a suppressed value premise, which is that U.S. leaders are motivated by the value of security from the Soviet military presence in the western hemisphere. If some other value had motivated policy-makers, either of the two competing causal explanations would have supported altogether different claims, for example, that the United States should invade and occupy Cuba.

Causal argumentation based on the D-N model of scientific explanation attempts to develop and test general propositions about the causes and effects of public policy.[27] Carl Hempel has elaborated D-N explanation as follows:

> We divide explanation into two major constituents, the *explanandum* and the *explanans*. By the *explanandum*, we understand the sentence describing the phenomenon to be explained (not that phenomenon itself); by the *explanans* the class of those sentences which are adduced to account for the phenomenon. . . . [Scientific explanation] answers the question, "Why did the *explanandum*-phenomenon occur?" by showing that the phenomenon resulted from particular circumstances, specified in $C_1, C_2, \ldots, C_k$, in accordance with laws $L_1, L_2, \ldots, L_r$. By pointing this out, the argument shows that, given the particular circumstances and the laws in question, the occurrence of the phenomenon was to be *expected;* and it is in this sense that the explanation enables us to understand why the phenomenon occurred.[28]

A simple example illustrates traditional causal (D-N) explanation.[29] If I leave my car outside overnight and the temperature drops below freezing, my full radiator (without antifreeze) will burst. Why will this happen? "My radiator burst" (*explanandum*). "My radiator was full of water, the cap was tightly fastened, and the temperature outside dropped below freezing" (circumstances, or C_k, in the *explanans*). And, "the volume of water expands when it freezes" (general proposition or law, L_r, in the *explanans*).[30] In this example, knowledge of prior circumstances and the appropriate law permits a prediction of the resultant event.

Questions have been raised about the suitability of D-N explanation in history and the social sciences.[31] These questions arise, among other reasons, because policy analysis and other social sciences are partly evaluative and advocative (normative) in character. Every advocative claim contains both factual and value premises, whereas in traditional causal explanations we evidently find only factual premises. Traditional causal explanations also

---

[25]Georg H. von Wright, *Explanation and Understanding* (Ithaca, NY: Cornell University Press, 1970), p. 145, and Kuhn, *Structure of Scientific Revolutions*.

[26]This is not to say that they do not imply values, because empirical theories in the social and natural sciences rest on unstated value premises. See, for example, M. Gunther and K. Reshaur, "Science and Values in Political 'Science,'" *Philosophy of Social Sciences* 1 (1971): 113–121; J. W. Sutherland, "Axiological Predicates in Scientific Enterprise," *General Systems* 19 (1974): 3–14; and Ian I. Mitroff, *The Subjective Side of Science: A Philosophical Inquiry into the Psychology of the Apollo Moon Scientists* (New York: American Elsevier Publishing, 1974).

[27]Allison, "Conceptual Models," p. 690 (note 4) tells us that Hempel's D-N (deductive-nomological) explanation is the basis (backing) for his three models.

[28]Hempel, Aspects of Scientific Explanation, pp. 247–258.

[29]See von Wright, *Explanation and Understanding*, p. 12.

[30]Ibid., p. 12.

[31]Ibid., p. 11.

require that the *explanans* precede (or accompany) the *explanandum*. Yet many advocative claims reverse this sequence, insofar as circumstances that explain action are situated in the future. Future circumstances, including intentions, goals, and desires, explain present actions to the extent that actions cannot occur without the motivation provided by such intentions, goals, and desires.[32] Finally, any correspondence between the results of acting on an advocative claim and the conclusions of a causal argument may be purely coincidental. In policymaking, predictions based on D-N explanations will fail if policy actors employ intelligent reflection to change their behavior, or if unpredictable factors deriving from creative thought intervene.[33]

D-N explanation is not the sole legitimate form of causal argumentation in public policy. Another form is hypothetico-deductive (H-D) explanation, which involves the deduction of hypotheses from theories that do not involve propositions about invariant causal relations or laws. Often the hypotheses of interest are those dealing with policy or program actions designed to achieve some practical outcome.[34] The relation between action and outcome, however, is not certain. If it were, it would conform to the following requirements, usually called the "essentialist" view of causation:

- The policy, *x,* must precede the outcome, *y,* in time.
- The occurrence of the policy, *x,* must be necessary for the occurrence of *y,* the outcome, which must not occur in the absence of the policy, *x*.
- The occurrence of the policy, *x,* must be sufficient for the occurrence of *y,* the outcome, which must occur when the policy, *x,* is present.

If these requirements were met, the relation between a policy action and an outcome would be certain. This requirement is virtually never satisfied in real-life policy settings. Instead, what occurs is that other conditions—uncontrolled contingencies that lie beyond the control of policymakers—make it impossible to know definitely whether a policy is necessary or sufficient for the occurrence of an outcome. The uncontrolled contingencies are plausible rival hypotheses that must be taken into account and, where possible, eliminated as competing explanations of a policy outcome. Here, the best that may be expected is an optimally plausible claim, that is, an approximately valid causal inference.[35] Figure 8.8 displays causal argumentation in the quasi-experimental tradition founded by Donald T. Campbell, a tradition based on the fundamental premise that causal argumentation in real-life policy settings requires the formulation, testing, and elimination of rival hypotheses.[36] Figure 8.8

---

[32]Ibid., pp. 74–124; G.E.M. Anscombe, *Intention* (London: Basil Blackwell, 1957), and W. H. Dray, *Laws and Explanation in History* (London: Oxford University Press, 1957).
[33]Alasdair Maclntyre, "Ideology, Social Science, and Revolution," *Comparative Politics* 5, 3 (1973): 334.
[34]See the discussion of the "activity theory of causation" in Thomas D. Cook and Donald T. Campbell, *Quasi-Experimentation: Design and Analysis Issues for Field Settings* (Boston, MA: Houghton Mifflin, 1979), ch. 1.
[35]The preeminent source on validity questions in the social and behavioral sciences is William R. Shadish, Thomas D. Cook, and Donald T. Campbell, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference* (Belmont, CA: Wadsworth, Cengage Learning, 2002).
[36]Rival hypotheses are also known as "threats to validity." See Donald T. Campbell, *Methodology and Epistemology for Social Science: Collected Papers,* ed. E. S. Overman (Chicago, IL: University of Chicago Press, 1988). The example is from Campbell's "Reforms as Experiments," in the Overman edition cited earlier.

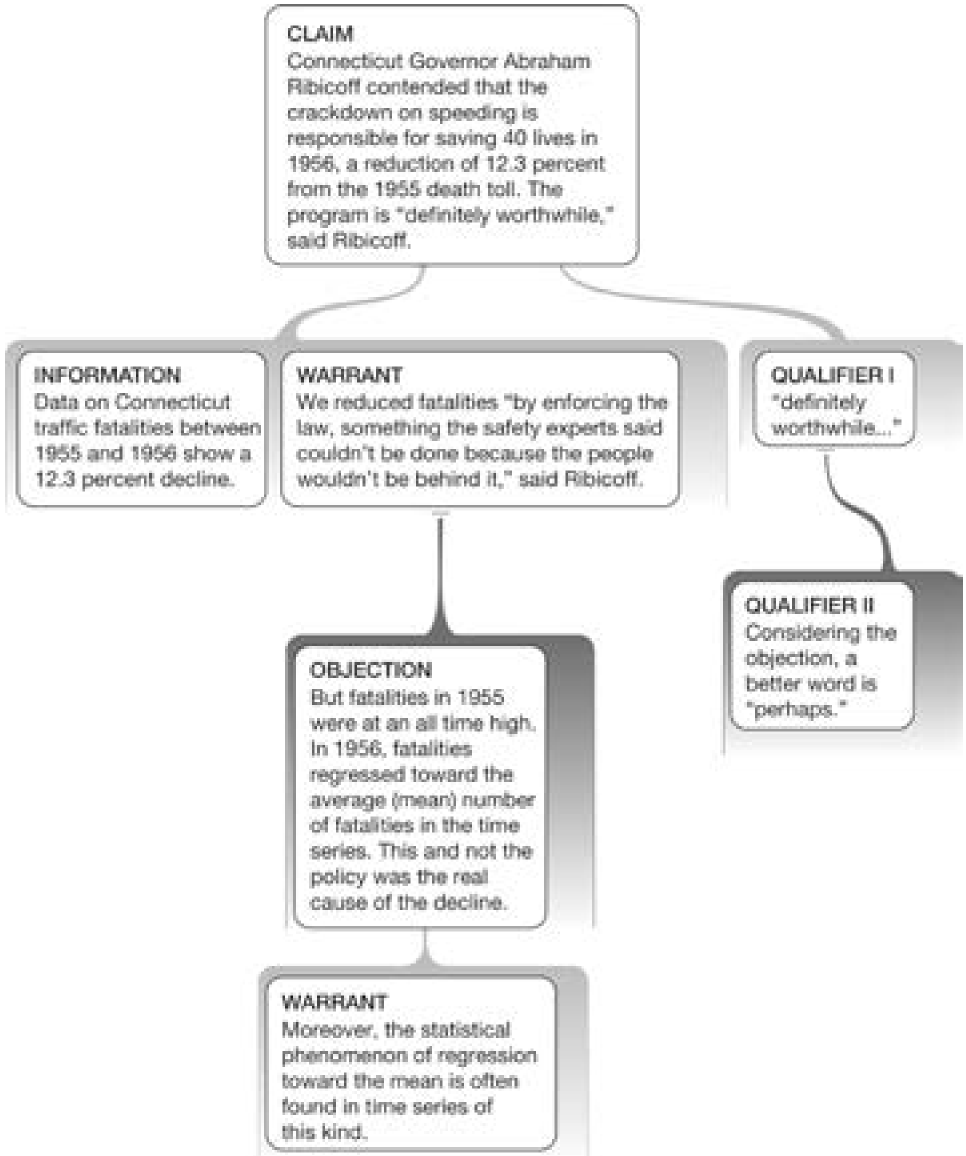


FIGURE 8.8
Argumentation from Practical Cause—Rival Explanations of the Effects of the Connecticut Crackdown on Speeding

shows that rival hypotheses used to challenge the warrant change the initial qualifier from "definitely worthwhile" (*Q1*) to "perhaps" (*Q2*).

Argumentation from Sign

Reasoning from sign is based on indicators and their referents. The presence of a sign indicates the presence of an event, condition, or process, because the sign and what it refers to occur together. Examples are indicators of institutional performance such as "organizational report cards," "benchmarks," and "best practices."[37] Another example is the widely used set of indicators (actually indices) of economic performance—"leading," "lagging," and "coincident" economic indicators—published periodically by the Conference Board. Signs are not causes. As we saw earlier, causality must satisfy additional requirements not expected of signs. Figure 8.9 displays an argumentation from sign based on a correlation coefficient and probability value (*p*-value).

Figure 8.9 makes an important distinction between signs and causes. In modern statistical analysis, measures of correlation, regression, and statistical significance (e.g., chi-square, *t*, and *F* values) are signs that refer to covariation. Covariation is a necessary but not sufficient condition of causation. In Figure 8.9, the *R* states a causal relation between participation in Head Start programs and increased graduation rates. It is a causal relation because students who participated had higher graduation rates; those who did not had lower graduation rates; the program came before the graduation in time; and there was, in addition, a positive correlation between participation and graduation. It is positive, moderate in strength ($r = 0.61$), and statistically significant ($p = 0.05$). The *O* successfully challenges the argument from sign, which mistakenly infers that Head Start is ineffective. The initial qualifier *Q1* ("probably") becomes *Q2* ("probably not").

Arguments from sign—whether presented as organizational report cards, benchmarks, or correlation coefficients—are at best statements of covariation or coincidence. Although covariation must be present for a causal relation to exist, causation requires that we fulfill certain conditions. John Stuart Mill, the nineteenth-century English philosopher, presented a set of methods (sometimes called "canons") of inductive inference designed to discover causal relations. *Mill's methods* are broadly employed today in the social and behavioral sciences, policy analysis, and program evaluation.[38] The methods are those of *agreement, difference, agreement and difference* (the so-called "joint method"), *concomitant variation,* and *residues*. The basic idea of the first three methods is as follows:

- If on two or more occasions a presumed effect has only one antecedent condition in common, then that condition is probably the cause of the presumed effect. If on the

[37]For example, William T. Gormley Jr. and David L. Weimer, *Organizational Report Cards* (Cambridge, MA: Harvard University Press, 1999).

[38]Quasi-experimental design in program evaluation is based very largely on Mill's methods. The best example is Cook and Campbell, *Quasi-Experimentation,* ch. 1, who critique and go beyond Mill's methods. For another critique and reformulation, see William N. Dunn, "Pragmatic Eliminative Induction," *Philosophica* 60, 2 (1997), Special issue honoring Donald T. Campbell. Comparative political science, comparative sociology, comparative public policy, and experimental psychology have also drawn on Mill's methods.

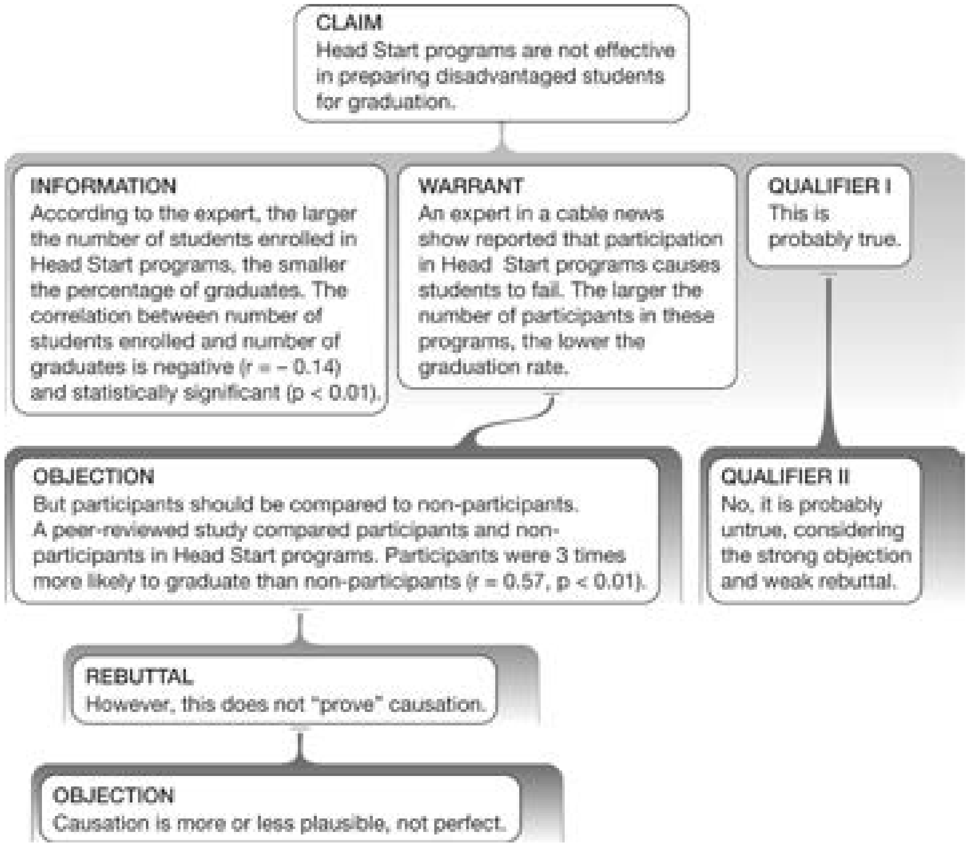


FIGURE 8.9
Argumentation from Sign—Quantitative Indicators Such as Correlation Coefficients and *P*-Values Do Not "Prove" Causation

first occasion, presumed effect Y is preceded by conditions X_1, X_3, and X_5, and on the second occasion the presumed effect Y is preceded by conditions X_2, X_3, and X_6, then X_3 is probably the cause of Y.

- If a presumed effect and a presumed noneffect share every antecedent condition except one, which occurs along with the presumed effect, then that condition is probably the cause of the presumed effect. If the presumed effect Y and the presumed noneffect Y share antecedent conditions X_1, X_2, X_5, and X_6, but do not share condition X_3, which occurs along with presumed effect Y, then X_3 is probably the cause of Y.
- If two or more occasions when a presumed effect occurs have only one antecedent in common, while two or more occasions when a presumed effect does not occur have nothing in common except the absence of that antecedent condition, then the antecedent condition in which the presumed effects and presumed noneffects differ

> is probably the cause. If on two or more occasions when presumed effect Y occurs it is accompanied solely by antecedent condition X_3, while two or more occasions when presumed effect Y does not occur have nothing in common except the absence of antecedent condition X_3, then X_3 is probably the cause of Y.

The method of concomitant variation, which we know today as correlation (covariation, association), does not require additional comment here, except to say that it is a necessary condition of causation. The method of residues is similar to what we now call the analysis of residual (error) variance in multivariate statistics and econometrics. The logic is that what is "left over" when we have explained the effects of all the (presumed) causes of a phenomenon are the "residues" of other possible (and usually unknown) causes. To know whether we are analyzing causation or correlation, however, requires that we first employ the first three methods, because no statistical analyses, however advanced, are sufficient to establish causation. Statistics provides signs, not causes.

Argumentation from Motivation

In motivational arguments, claims assert that an action should be adopted because of the motivating power of intentions, goals, or values. Motivational arguments seek to demonstrate that the intentions, goals, or values underlying a recommended course of action are such as to warrant its acceptance, adoption, or performance. It is often sufficient to know that large or important groups actually desire to follow the course of action stated in the claim.

Argumentation from motivation represents a form of reasoning that philosophers since Aristotle have called the *practical syllogism*, or *practical inference*. In practical inference, the major premise or warrant W describes some desired state or end of action, while the minor premise or information I relates a course of action to this desired state as a means to an end. The conclusion or claim C consists of a recommendation act in a certain way to secure the desired state or end. Whereas in *theoretical inference* (argumentation from cause) acceptance of a general proposition or law leads to the conclusion or claim, in *practical inference* (argumentation from motivation) acceptance of a premise about goals, values, or intentions leads a conclusion or claim about actions that are in accordance with them.[39] Claims in practical inference are usually designed to *understand actions,* whereas claims in theoretical inference seek to *explain events*.[40]

Practical reasoning is of great importance to policy analysis, where one of the chief problems is to explain actions in terms of goals, values, and intentions:

> the practical syllogism provides the sciences of man with something long missing from their methodology: an explanation model in its own right which is a definite

[39]von Wright, *Explanation and Understanding*, pp. 22–27.

[40]Ibid., pp. 22–24. See also Fred R. Dallmayr and Thomas A. McCarthy, eds., *Understanding and Social Inquiry* (Notre Dame, IN: University of Notre Dame Press, 1977); and Rein, *Social Science and Public Policy,* pp. 14–15, 139–170.

> alternative to the subsumption-theoretic covering law model [i.e., deductive-nomological explanation—*author*]. Broadly speaking, what the subsumption-theoretical model is to causal explanation and explanation in the natural sciences, the practical syllogism is to teleological explanation and explanation in history and the social sciences.[41]

Motivational argumentation not only provides an alternative explanatory model for policy analysis, but also compels us to conceptualize policymaking as a political process. Arguments from motivation force analysts to think in terms of the goals, values, and intentions of policy actors, and "enter the phenomenological world of the policy maker."[42] Motivational arguments also bring us closer to questions of values and ethics, which in other modes of policy argumentation are frequently and regrettably separated from "factual" matters. Figure 8.10 presents an argument from motivation that is challenged by another argument from motivation and an argument from classification. In this example, *Q1* ("definitely") becomes *Q2* ("probably") after the rebuttal.

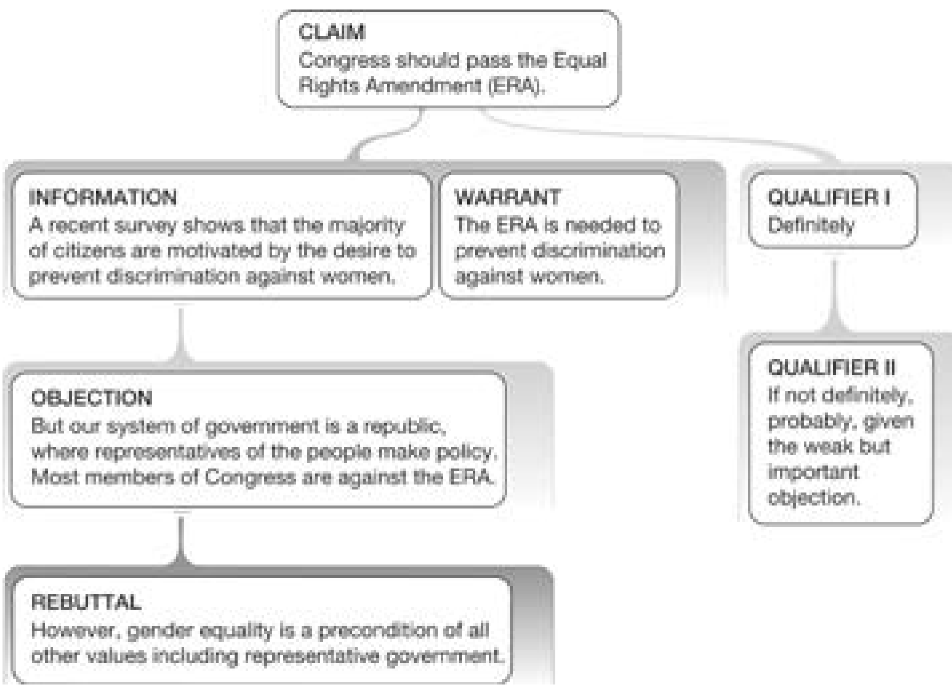


FIGURE 8.10
Argumentation from Motivation—Support for the Equal Rights Amendment

---

[41]von Wright, *Explanation and Understanding*, p. 27.
[42]Bauer, *Study of Policy Formation*, p. 4.

Argumentation from Intuition

In reasoning from intuition, policy claims are based on premises or assumptions about the insight of participants in the policymaking process. Policy-relevant information consists of factual reports or expressions opinion. The function of the warrant is to affirm that inner mental states (insight, judgment, understanding) of producers of information make them specially qualified to offer opinions or advice. The policy claim may simply reiterate the report or opinion supplied in the information. Consider this example of early military policy:

> When in 1334 the Duchess of Tyrol, Margareta Maultasch, encircled the castle of Hochosterwitz in the province of Carinthia, she knew only too well that the fortress, situated on an incredibly steep rock rising high above the valley floor, was impregnable to direct attack and would yield only to a long siege. In due course, the situation of the defenders became critical: they were down to their last ox and had only two bags of barley corn left. Margareta's situation was becoming equally pressing, albeit for different reasons: her troops were beginning to be unruly, there seemed to be no end to the siege in sight, and she had similarly urgent military business elsewhere. At this point the commandant of the castle decided on a desperate course of action, which to his men must have seemed sheer folly; he had the last ox slaughtered, had its abdominal cavity filled with the remaining barley, and ordered the carcass thrown down the steep cliff onto a meadow in front of the enemy camp. Upon receiving this scornful message from above, the discouraged duchess abandoned the siege and moved on.[43]

Figure 8.11 serves to emphasize some of the unique advantages of insight, judgment, and tacit knowledge in developing creative solutions to policy problems. However, it also points to difficulties. Although policy scholars urge that intuition, judgment, and tacit knowledge be incorporated into policy analysis,[44] it is seldom possible to identify in advance the methods or forms of reasoning that are likely to yield insight or creativity. A creative act, observes Churchman, "is an act that cannot be designed beforehand, although it may be analyzable in retrospect. If this is the correct meaning of creativity, then no intelligent technician can be creative."[45]

Argumentation from Analogy

Reasoning from analogies and metaphors is based on similarities between relationships found in a given case and relationships found in a metaphor, analogy, or allegory

[43]Quoted in Paul Watzlawick, John Weakland, and Richard Fisch, *Change: Principles of Problem Formation and Problem Resolution* (New York: W. W. Norton & Company, 1974), p. xi.

[44]For example, Yehezkel Dror, *Ventures in Policy Sciences* (New York: American Elsevier Publishing, 1971), p. 52; Sir Geoffrey Vickers, *The Art of Judgment: A Study of Policy Making* (New York: Basic Books, 1965); and Edgar S. Quade, *Analysis for Public Decisions* (New York: American Elsevier, 1975), pp. 4–5. Some observers have also commented favorably on the possibility of drug-induced changes in the mental states of policymakers. See Kenneth B. Clark, "The Pathos of Power: A Psychological Perspective," *American Psychologist* 26, 12 (1971): 1047–1057.

[45]Churchman, *The Design of Inquiring Systems*, p. 17.

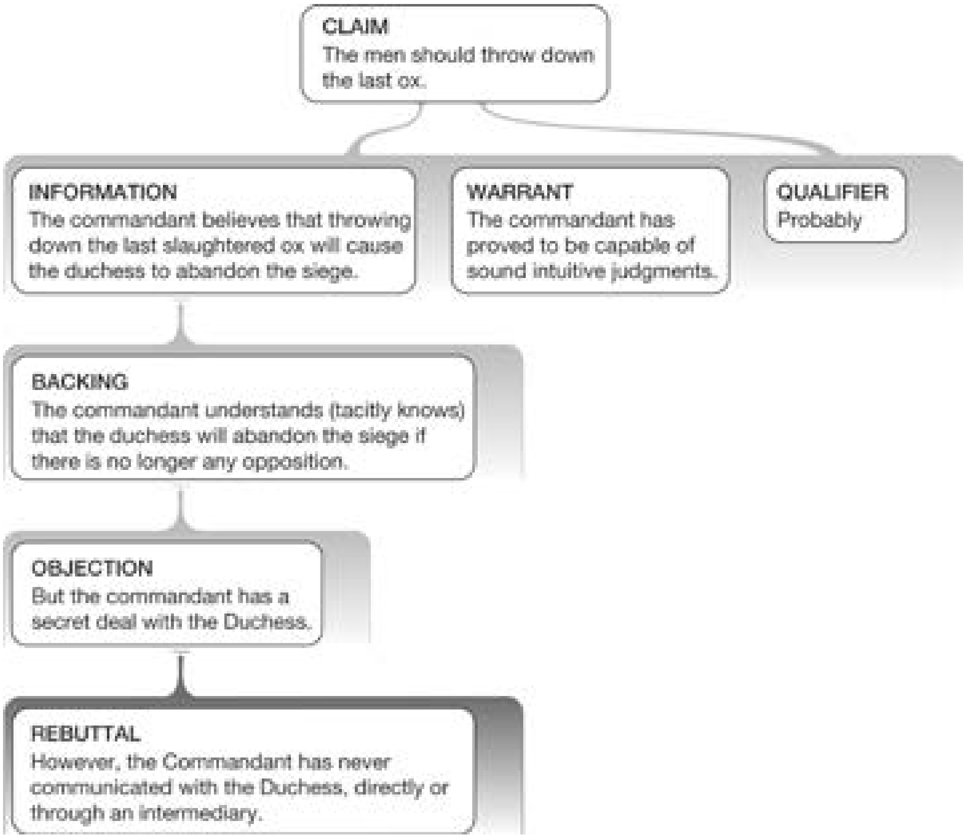


FIGURE 8.11
Argumentation from Intuition—A Counterintuitive Solution Avoids Certain Death

(Figure 8.12). For example, the claim that government should "quarantine" a country by interdicting illegal drugs—with the illegal drugs presented as an "infectious disease"—is based on reasoning that, because quarantine has been effective in cases of infectious diseases, interdiction will be effective in the case of illegal drugs. In arguments from analogy, claims are based on assumptions that relationships among two or more cases (not the cases themselves) are essentially similar. For example, claims about the desirability of adopting policies to mitigate air pollution have been based on beliefs about the success of water pollution policies. In making claims about ways to reduce employment discrimination against women, the reasoning is sometimes based on assumptions about the success or failure of policies designed to reduce discrimination against ethnic minorities.[46]

[46]Robert L. Lineberry, *American Public Policy: What Government Does and What Difference It Makes* (New York: Harper and Row, 1977), p. 28.

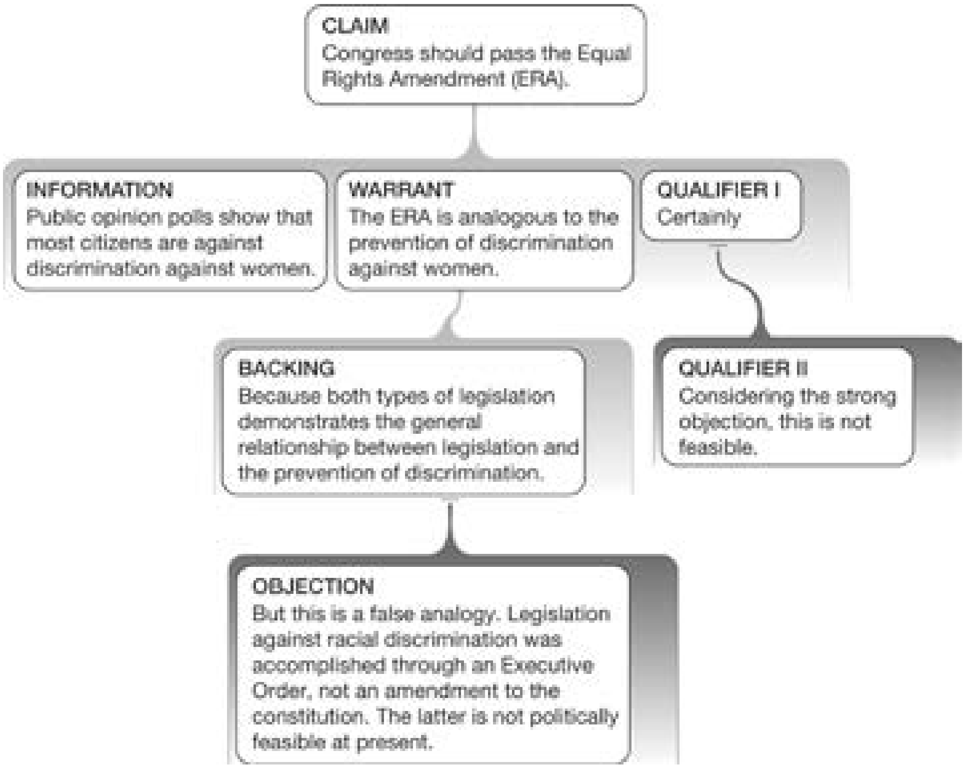


FIGURE 8.12
Argumentation from Analogy—The Equal Rights Amendment and False Analogy

Argumentation from Parallel Case

Reasoning from parallel case focuses on similarities among two or more (Figure 8.13) cases of policymaking. For example, a reason that a local government should strictly enforce pollution standards is that a parallel policy was successfully implemented in a similar local government elsewhere. Policy claims are based on assumptions that the results of policies adopted in similar circumstances are worthwhile or successful. Government agencies in the United States and abroad often face similar problems, and policy claims may be based on their common experiences. The British experiences with comprehensive medical care and city planning ("new towns"), and the Dutch and Swiss approaches to the decriminalization of illegal drugs, have influenced debates about drug policy in the United States. The experience of some states in adopting taxation, open housing, and equal employment opportunity policies has been used as a basis for policy at the federal level.[47] A variation of argumentation from parallel case is an argument based on the experience of the same agency over time.

[47] Ibid.

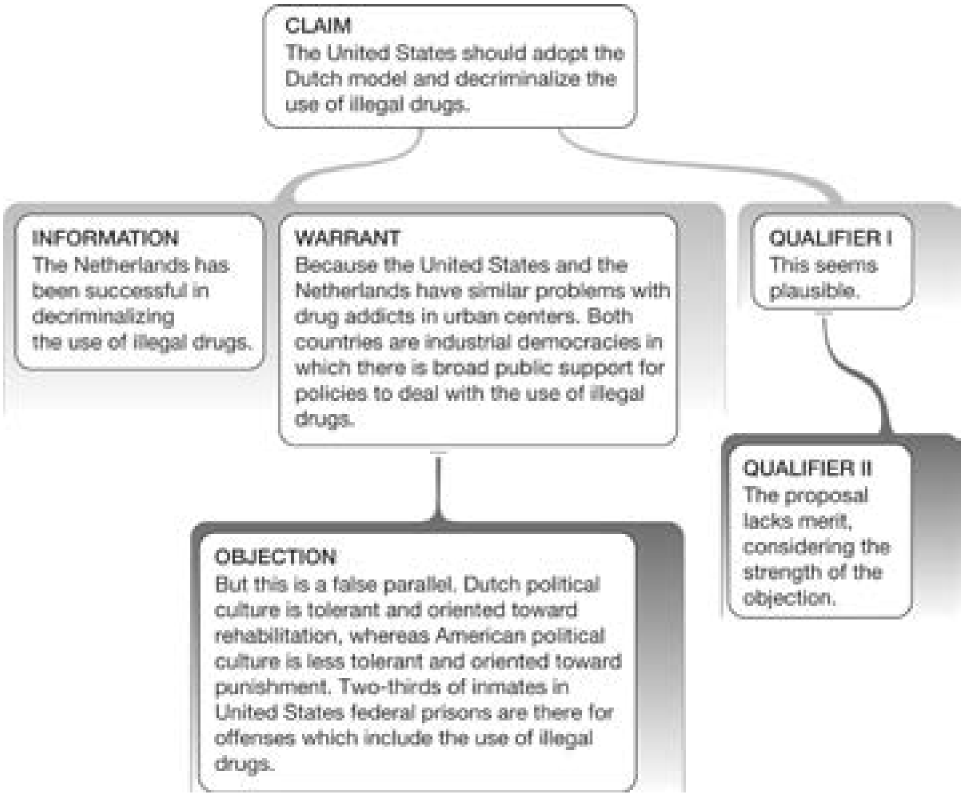


FIGURE 8.13
Argumentation from Parallel Case—The Dutch Model and False Parallel

Past policies in the same agency are used to support claims that the agency should adopt particular courses of action, usually those that are marginally different from the status quo. Claims about federal and state budgetary policies are typically based on assumptions about similarities with past policies adopted in the same agency.[48]

Argumentation from Ethics

Reasoning from ethics is based on the rightness or wrongness, goodness or badness, of policies or their consequences (Figure 8.14). For example, policy claims are frequently based on moral principles stating the conditions of a "just" or "good" society, or on ethical norms prohibiting lying in public life. Moral principles and ethical norms go beyond the values

[48]See, for example, Aaron Wildavsky's now classic treatment of incremental policymaking in *The Politics of the Budgetary Process* (Boston: Little, Brown and Company, 1964). See also the discussion of models of policy change in Chapter 2.

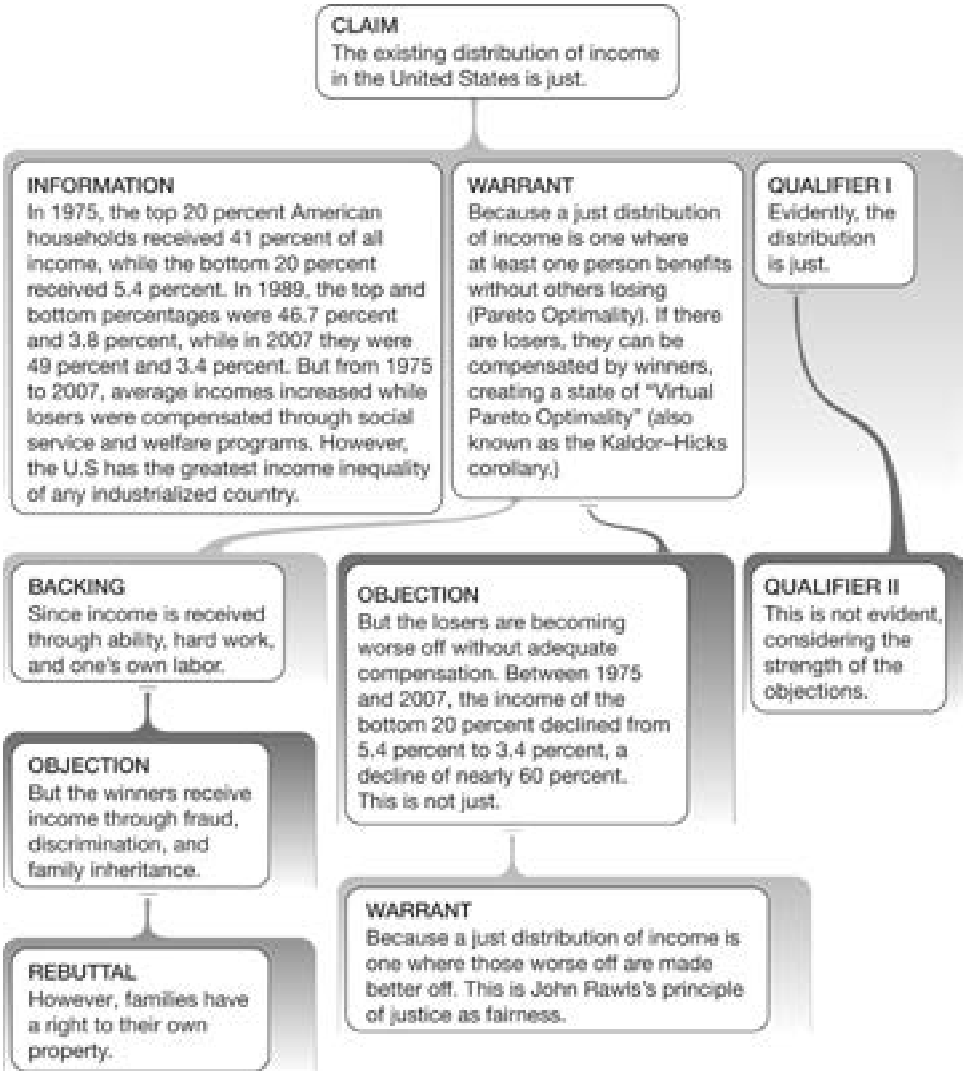


FIGURE 8.14
Argumentation from Ethics—Income Distribution and Justice as Fairness

and norms of particular individuals or groups. In public policy, many arguments about economic benefits and costs involve unstated or implicit moral and ethical reasoning. The warrant in an argument from ethics provides reasons for accepting a claim by associating it with some moral principle or ethical rule. The claim is that the person, situation, or condition referred to in the information should be regarded as valuable or worthless or that a policy described in the information should or should not be adopted.

To illustrate ethical argumentation, consider Figure 8.14. Here, the evaluative claim, *C*, is that "the existing distribution of income in the United States is just." The information, *I*, is:

> In 1975, the top 20 percent of American families received 41 percent of all income, while the bottom 20 percent received 5.4 percent. In 1989, the top and bottom percentages were 46.0 and 3.8 percent, while in 2007 they were 49.0 and 3.4 percent. From 1975 to 2007, while average incomes increased by about 3 percent, winners compensated losers through social service and welfare programs. The United States has the greatest income equality of any industrialized state.

The warrant, *W*, is the *Pareto rule,* named after the Italian economist and sociologist Vilfredo Pareto (1848–1923).

The Pareto rule is a simple ethical principle supported by many policy economists.[49] The rule states that "an optimum distribution of income in society is one where some individuals benefit without others losing." The backing, *B*, for the warrant is "Since income is received through ability, hard work, and one's own labor." The objection, *O*, is "But the winners receive income through fraud, discrimination, and inheritance." The rebuttal, *R*, to the objection is "However, families have a right to their own property" (Figure 8.14).

Figure 8.14 also shows how an important contending moral principle, John Rawls's principle of "justice as fairness," serves as an objection to the Pareto principle. The example of Pareto optimality also demonstrates that a widely accepted ethical rule, while it supports claims about a just society, does not apply to situations involving fraud, discrimination, and inheritance. By engaging in the systematic analysis of underlying ethical and moral reasoning, parties to a debate are compelled to clarify the meaning of key concepts such as "entitlement," which are more complex than may be apparent at first glance. Parties making a claim also may be compelled to consider whether particular ethical rules, such as Pareto and Rawlsian optimality, violate their own moral convictions. Proponents of the Pareto Rule may see that its application violates moral convictions about the necessity of basing principles of entitlement on ability and work. If this is done for welfare recipients, why not for heirs to estates? In short, the analysis of ethical argumentation can help us probe ethical rules and moral principles to determine whether they are general in their applicability and internally consistent.[50] Ethical argumentation, it should be emphasized, differs from each of the other modes of reasoning in one essential respect: Whereas each of the other modes takes values as "given"—for example, values which are described in public opinion surveys are a basis for arguments from motivation (Figure 8.10)—the process of ethical argumentation attempts to discover whether there are *good reasons* to make ethical claims.

EVALUATING POLICY ARGUMENTS

The evaluation of policy arguments facilitates critical thinking in public policy analysis. So far, we have looked at the structure of arguments, the process of argumentation, and modes of reasoning employed in making policy claims. This enables the identification of hidden or tacit

[49]See Peter G. Brown, "Ethics and Policy Research," *Policy Analysis* 2 (1976): 332–335.

[50]MacRae, *Social Function of Social Science*, pp. 92–94.

assumptions and shows how the plausibility of a claim, as expressed in its qualifier, changes as a result of objections and rebuttals introduced in the course of policy argumentation.

We now turn to specific criteria for evaluating arguments. Some of these criteria come from propositional logic, a discipline that offers criteria for determining—without qualification and with deductive certainty—the *formal* validity of arguments.[51] Other criteria originate in a continuously evolving body of standards for assessing the *informal* validity of arguments.[52] Still other criteria come from procedures employed by methodologists who work in the hermeneutics and interpretationist (*Verstehende*) traditions. These traditions have long been concerned with discovering and accurately representing the meaning of human actions, whether expressed in the form of written texts or as spoken language.[53] Finally, some criteria originate in philosophical pragmatism, an epistemology and methodology that is useful for evaluating entire systems of argumentation.[54]

Some Hermeneutic Guidelines

Hermeneutics investigates the meanings of human texts. It is perhaps the most comprehensive and systematic of the qualitative methodologies. The term *qualitative*, contrary to a common misunderstanding in the social sciences,[55] is not simply the logical complement of "quantitative," in the sense that qualitative refers to that which is *not* quantitative or statistical. To be sure, it is inappropriate to employ the term "qualitative" to small-*n* case study research or the discrete (versus continuous) level of measurement. Rather, qualitative methods—also known as hermeneutic, interpretive, ethnographic, or constructivist methods—are expressly designed to investigate the meanings of individual and collective action.[56] Statistical and other forms of quantitative analysis are not appropriate for this purpose.

Human texts refer not only to written documents that are products of human action, for example, legislative transcripts and laws that originate as policy debates; they also refer to the actions themselves, whether or not they have been expressed in written form. Among various ways to demonstrate the importance of hermeneutic guidelines to the evaluation of policy arguments, a focus on ethical argumentation is most revealing. Consider alternative ethical arguments about a prisoner captured by soldiers from the opposing army:

- *Argument A.* The prisoner is not a member of a regular fighting force. He is a common criminal, a terrorist who lies when interrogated about terrorist plots and other military secrets. Because he does not qualify formally as a "prisoner of war," he should not be protected against inhumane treatment and torture.

[51]Classic sources on formal logic include Irving M. Copi, *An Introduction to Logic* (New York: Macmillan, 1953).
[52]See Toulmin, Rieke, and Janik, *An Introduction to Reasoning*, Part V: Fallacies: How Arguments Go Wrong, pp. 129–197.
[53]The classic treatise on hermeneutics, originally published in 1960, is Hans Georg Gadamer, *Truth and Method* (New York: Seabury, 1975).
[54]See Nicholas Rescher, *Induction* (Pittsburgh, PA: University of Pittsburgh Press, 1980).
[55]See, for example, a widely used text by Gary King, Robert Keohane, and Sidney Verba, *Designing Social Inquiry: The Logic of Inference in Qualitative Research* (Princeton, NJ: Princeton University Press, 1994).
[56]For an extended critical essay on the meaning of qualitative see John Searle, *The Construction of Social Reality* (New York: The Free Press, 1995).

- *Argument B.* The prisoner is a freedom fighter who is combating the terror and oppression inflicted by the enemy. He is a loyal soldier who has a moral duty to mislead interrogators about military secrets. He should be protected as a prisoner of war.
- *Argument C.* The prisoner is a member of a fighting force constituted by a standing army. He is a prisoner of war whose loyalty demands that he mislead interrogators about military secrets. As such, he should receive the protections against inhumane treatment afforded any prisoner of war.

The evaluation of these arguments can benefit from guidelines for the interpretation of written or oral argumentation (Exhibit 8.2).

EXHIBIT 8.2

Guidelines for Interpreting Arguments

- Policy argumentation has three major functions: to generate debate that improves the validity, soundness, and efficacy of policies (dialectical function); to present optimally valid and empirically sound conclusions (logical-empirical function); to persuade others to accept policy arguments (rhetorical function), apart from the validity, soundness, or usefulness of the arguments.
- Look for concealed meanings in words, sentences, and entire arguments. A word or sentence may not mean what it says on the surface. Example: "He is a good Liberal" does not mean that the person described performs well as a Liberal; it rather means that the person's identity as a Liberal is associated with some kind of weakness or limitation.
- Distinguish between the surface meaning of a word, sentence, or argument and its meaning in the context of the arguer. Try to identify any differences in your understanding from that of the arguer. Example: "The mayor should not have publicly acquiesced in the demonstrators' demands." Several potential misinterpretations are possible: someone except the mayor should have acquiesced; the mayor should not have acquiesced in public; the mayor should not have acquiesced at all.
- Observe the principle of hermeneutic charity, which requires that discrepancies in meaning be resolved by accepting, or trying to understand, what the arguer is trying to say. Example: Critics of arguments presented in quantitative language often label such arguments (and the arguers) as "logical positivists," notwithstanding the fact that quantification, per se, has nothing to do with logical positivism. A charitable effort to understand what the arguer actually believes can solve this problem.
- Look for terms that are used pejoratively to discredit a person or policy. On the surface, these terms can be neutral; but in context, they are often used pejoratively. Examples: "This is just another example of a new bureaucracy." "These are the arguments of typical 'tree-huggers.'" "The report, written by a bunch of logical-positivists, is unacceptable."

The main function of arguments A and B is rhetorical, not dialectical or logical-empirical. Although the use of A and B as rebuttals to one another would contribute to the dialectical function, this is not the purpose of A or B alone. A and B do present evidently factual statements, some of which appear to be empirically sound and even uncontroversial. A prisoner was captured, he was involved as a combatant, and he was not telling the truth. What is disputed are two main issues: Is it is morally and legally right to lie under the circumstances? Does the soldier qualify as a prisoner of war?

Other guidelines also apply. The quotation marks around some of the words ("prisoner of war" and "inhumane treatment") conceal or obscure meanings. The contexts of A and B are also important for understanding the arguments. Given the history of many conflicts, it is not surprising that both sides regard opponents as terrorists. The principle of hermeneutic charity encourages giving the benefit of the doubt to each party, for example, by trying to understand that both seek to be treated equally. The one side contests the moral and legal acceptability of attacks on their troops (and on civilians) by out-of-uniform combatants, arguing that such practices are unfair. The other side affirms the moral and legal acceptability of such acts, on grounds that such acts are fair under conditions where one side has a preponderance of modern weaponry ("asymmetric warfare"). Finally, both arguments make liberal use of pejoratives—"extremist," "criminal," "terrorist," "propagandists," "freedom fighter," "oppression"—that obscure rather than clarify moral and legal issues.

If we were to rephrase the arguments, introducing rebuttals and replacing absolute with qualified claims, it would look something like argument C. Argument C has the advantage of isolating the contested issue, which is one of *conflicting obligation:* In the ethics of war, there has been a widespread understanding that prisoners who mislead interrogators about military secrets are displaying courage, honor, patriotism, and other virtues. Here, the same understanding would presumably apply to both sides.

Guidelines from Informal and Formal Logic

Hermeneutic guidelines are designed to enhance the understanding of meanings that underlie arguments and argumentation. Questions concerning the soundness, credibility, or plausibility of arguments do not arise, because the principal aim is to achieve an accurate interpretation of what arguers mean. By contrast, the fields of informal and formal logic provide guidelines for recognizing and assessing the significance of informal fallacies.[57] Here as elsewhere the term "guideline" is used in place of "rule," because there is no way to determine absolutely whether an argument is fallacious. Hence, the analysis of informal fallacies does not permit all-or-none conclusions.

As we saw in the first part of this chapter, there are numerous modes of argumentation that are generally recognized as appropriate to policy discourse. Arguments of the following kinds are *formally* valid:

- *Hypothetical syllogism.* If p implies q, and q implies r, then p implies r. Or: p q, q r, p r (= implies). Example: A transitive preference ordering is one form of

[57]The terms "informal logic" and "informal fallacy" are used in the discipline of logic, where "formal logic" and "formal fallacy" are also distinguished.

the hypothetical syllogism. Given three projects, *A*, *B*, and *C*, if A *p* B, and B *p* C, then A *p* C (*p* = preferred to). This formally valid argument can be empirically unsound.

- *Modus ponens.* *Modus ponens* (method of affirming) asserts that if *p* implies *q*, and *p* occurs, then *q* will occur. If *p* *q*, and *p*, then *q*. Example: If investment *I* in a project produces outcome *O*, and investment *I* is made, then outcome O will be the result. Although this argument is formally valid, the conclusion assumes that no causally relevant factor other than *I* is present, a situation that almost never exists. This formally valid argument can be empirically unsound.
- *Modus tollens.* *Modus tollens* (method of denying) asserts that if *p* implies *q*, and *q* does not occur, then *p* will not occur. If *p* *q*, and *q* does not occur (~ *q*), then *p* will not occur (~ *p*). Example: If investment *I* in a program produces outcome *O*, and *O* does not occur, then *I* is not a cause. This formally valid argument can be empirically unsound.

We now turn to modes of argumentation that are generally recognized as formally invalid, inappropriate, or unsound—however persuasive they may appear on first glance. These modes of argumentation are called fallacies. A fallacy is an argument that is weakened or seriously flawed because it uses irrelevant or inadequate information, erroneous or unsound reasoning, or inappropriate and misleading language. Table 8.2 provides a listing of fallacies and guidelines that are helpful in recognizing them.

TABLE 8.2

Guidelines for Identifying Invalid Arguments and Fallacies

Fallacy	Guideline
Affirming the Consequent	A matter of formal (propositional) logic. A logically invalid argument is: If *p* then *q*, and *q*, then *p* (*p* *q*, *q*, therefore *p*). "The paradigm of 'proof through prediction,'" notes Merton, "is, of course, logically fallacious: If A (hypothesis), then B (prediction). B is observed. Therefore, A is true." * If the strict enforcement of speeding causes traffic accidents to decline, and traffic accidents decline, then strict enforcement was the cause. In scientific research, this form of argument, although formally invalid, can be useful because a hypothesis may be improved by testing conditions other than B, that is, rival hypotheses.
Denying the Antecedent	Again, a matter of formal logical validity. If *p* then *q*, and not-*p*, then not-*q* (*p* *q*, ~ *p*, therefore ~ *q*) is fallacious. Example: Because market economies are democracies, and country X is not a market economy, it is not a democracy.
False Analogy	The comparison of two relationships believed to be similar disregards important differences that make the comparison relatively unsound. Example: Because drug addiction is like an infectious disease, quarantining addicts is the only policy that will work.
False Parallel	The comparison of two cases believed to be similar disregards important differences that make the comparison unsound. The acquiescence of the United States in World War II led to genocide and ethnic cleansing. The United States cannot acquiesce in ethnic cleansing in the Balkans.

Fallacy	Guideline
Hasty Generalization	In making a generalization from particular instances of a case, a failure to recognize that there are too few instances of the case, or that the instances are exceptional rather than typical. In conducting opinion surveys, an inadequate sample size will yield too few instances, while a failure to use random sampling—where every element or instance has an equal chance of being selected—is likely to yield exceptional conclusions rather than those typical of the population. Example: "Focus group" interviews with fifteen typical voters conducted before the election show that there is greater support for candidate A than candidate B.
False Cause	In making a claim about cause and effect, arguing that a single cause is responsible for an effect, but without examining other plausible causes. False causes also stem from confusing statistical correlation or covariance with causality and inferring cause from temporal sequence alone (post hoc fallacy). Examples: Excessive government spending is responsible for the slow growth of GDP (single false cause). That economic conditions affect social well-being is evident from the statistically significant positive correlation ($r = 0.74$, $p = 0.05$) between suicide and unemployment (false cause based on correlation). After the Reagan (or Clinton) administration took office, we had the highest unemployment (or government spending) in 20 years (post hoc fallacy).
Fallacy of Composition	Concluding that something is true of the whole because it is true of its parts. The fallacy of composition (also called the aggregative or holistic fallacy) involves all parts, not just a sample, so it differs from the fallacy of hasty generalization (see earlier). Example: Vehicle safety studies of the severity of damage suffered by test robots riding at different speeds in automobiles show that speed and severity of damage are strongly and positively correlated. This is striking evidence that "speed kills!" But studies of fatal accidents show that approximately 20 percent of fatal accidents are related to speeding.
Fallacy of Division	Concluding that something is true of the parts because it is true of the whole (also called the individualistic fallacy). Example: Because the per capita income of a country has increased, everyone is better off. In many countries, however, this is false. Persons who are better off become even better off, while those worse off become even worse off. Another example is using the arithmetic mean and other averages to describe a group, without examining differences among the group's members (e.g., outliers in a scatterplot).
Fallacy of the Slippery Slope	Concluding on the basis of insufficient or inadequate evidence that if one event occurs, then others will follow in an inevitable or uncontrollable sequence. Example: If the legislature passes a new law requiring stricter registration of handguns, it will lead to government confiscation of all guns.
Begging the Question	A claim is assumed as a reason or evidence. Example: "With a force of 500,000 troops we will be able to invade Iraq and topple President Saddam Hussein. It will take this many troops to do the job."
Ad Hominem	An individual's personal characteristics are used as part of an argument, when such characteristics are irrelevant to an issue. Examples: "An eminent natural scientist concludes that welfare reform is unsound." "The argument of the environmentalists is deeply flawed. After all, these 'tree huggers' are just socialists in disguise." "Theories of economic

(*continued*)

Fallacy	Guideline
	development are products of Western thinking. Obviously, they are inapplicable to the non-Western world." Note that when personal characteristics are relevant, no ad hominem fallacy is involved. For example, "expert witnesses" in court cases should have appropriate expertise.
Ad Populum	The characteristics or beliefs of a group or community are used as part of an argument, when the characteristics or beliefs are irrelevant to the issue. "The majority of the community believes that fluoride causes cancer."
Appeal to Tradition	A claim is based on conformity to tradition, when tradition is largely or entirely irrelevant to the issue. Examples: "We have always done it this way." "The founding fathers would be appalled by the Senator's proposal." "The successful disciplines have succeeded because they have emulated physics. This should be the model for the social sciences, if they want to succeed."
Accent	A misplaced emphasis on a word, phrase, or portion of an argument results in misunderstanding or misinterpretation. The use of italics, boldface print, variable fonts, photos, clip-art, and colors can accentuate the importance of relatively sound or plausible, as well as relatively unsound or implausible arguments, or parts of arguments. A leading example of the fallacy of accent is quoting or extracting information, reasons, or arguments out of context.

Note: * Robert K. Merton, *Social Theory and Social Structure,* rev. ed. (Glencoe, IL: Free Press, 1957), p. 99n. The same point is made by Donald T. Campbell, *Epistemology and Methodology for Social Science: Selected Essays* (Chicago, IL: University of Chicago Press, 1988), p. 168.

CHAPTER SUMMARY

This chapter has provided an understanding of the structure and process of policy argumentation, focusing on contrasts among types of claims, the identification and arrangement of elements of policy arguments, and the effects of rebuttals on the dynamics of argumentation. The chapter also contrasts different modes of policy reasoning and offers guidelines for the identification and assessment of common fallacies that weaken or seriously flaw policy arguments. Policy argumentation is central to policy analysis and the policy-making process.

REVIEW QUESTIONS

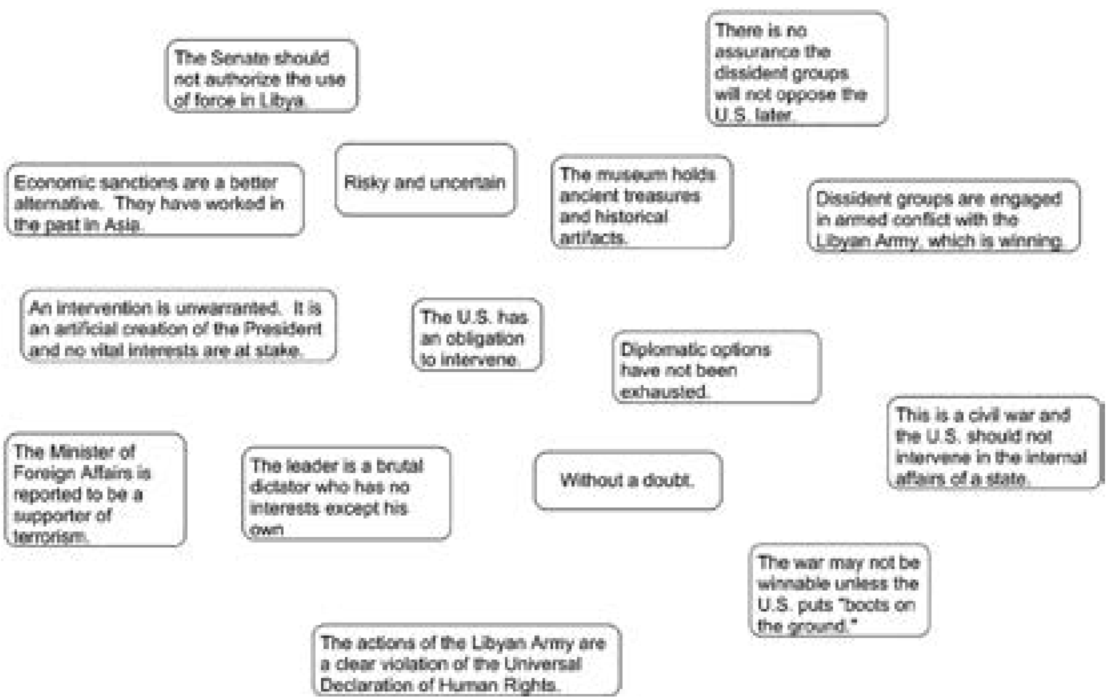
1. Use three of the following terms to construct definitive, designative, evaluative, and advocative claims. There should be twelve claims in all.

crime	fiscal crisis
pollution	human rights
terrorism	ethnic cleansing
quality of life	Unemployment
global warming	Poverty

2. Develop a policy argument on the basis of one or more of the terms in question 1.
3. Convert the argument in question 2 into a policy debate by providing an objection and a rebuttal.
4. Explain why the qualifier changed (if it did) after introducing the objection and rebuttal. If the qualifier did not change, why?
5. Define the term "fallacy?" Does the commission of a formal fallacy such as *affirming the consequent* invalidate an argument? Is the same true for informal fallacies, for example, *false analogy* or *ad populem*?

DEMONSTRATION EXERCISES

1. Obtain an online or hard copy of the international affairs section of a newspaper. Identify and describe as many *modes of argument* as you can. Would you expect to find different modes of argument in academic journals than in newspapers? Explain.
2. Read the letters to the editor in a newspaper, magazine, or online bulletin board or blog. Find as many examples of formal and informal fallacies as you can. A variation of this exercise is to break into groups to complete the assignment.
3. Read Case 8.1 (Pros and Cons of Balkan Intervention), which is drawn from an editorial in the *Los Angeles Times*. Use the argument mapping procedures presented in this chapter to analyze the pros and cons (or strengths and weaknesses) of the recommendation that the United States should *not* intervene in the Balkans. In doing this exercise, either display the elements of argument with Microsoft Draw or use Rationale 2, the special computer program for mapping the structure of policy arguments.
4. Write a one-page analysis in which you assess the overall plausibility of the claim "the conflict in Bosnia is somebody else's trouble. The United States should not intervene militarily." Hand in a printed copy of your argument map.
5. Below is an argument map in which the warrants, backings, objections, rebuttals, and qualifiers have been scrambled.[58] Rearrange the elements to make a persuasive argument and counter-argument. Study Case 8.2 as an example.



[58]This may be done with a utility in *Rationale 2*.

BIBLIOGRAPHY

Alker, Hayward R., Jr. "The Dialectical Logic of Thucydides' Melian Dialogue." *American Political Science Review* 82, 3 (1988): 805–820.

Anderson, Charles W. "Political Philosophy, Practical Reason, and Policy Analysis." In *Confronting Values in Policy Analysis*. Edited by F. Fischer and J. Forester. Newbury Park, CA: Sage Publications, 1987.

Apthorpe, Raymond, and Des Gasper, eds. *Arguing Development Policy: Frames and Discourses*. London: Frank Cass, 1996.

Campbell, Donald T. *Epistemology and Methodology for Social Science: Selected Essays*. Chicago, IL: University of Chicago Press, 1988.

Dunn, William N. "Policy Reforms as Arguments." In *The Argumentative Turn in Policy Analysis and Planning*. Edited by F. Fischer and J. Forester. Durham, NC: Duke University Press, 1993.

Fischer, David H. *Historians' Fallacies: Toward a Logic of Historical Thought*. New York: Random House, 1970.

Fischer, Frank, and John Forester, eds. *The Argumentative Turn in Policy Analysis and Planning*. Durham, NC: Duke University Press, 1993.

Gasper, Des, and R. V. George. "Analyzing Argumentation in Planning and Public Policy: Improving and Transcending the Toulmin-Dunn Model." *Environment and Planning B: Planning and Design* 25 (1998): 367–390.

Majone, Giandomenico. *Evidence, Argument, and Persuasion in the Policy Process*. New Haven, CT: Yale University Press, 1989.

McCloskey, Deirdre N. *The Rhetoric of Economics*. Madison: University of Wisconsin Press, 1988.

Merton, Robert K. *Social Theory and Social Structure*. Rev. ed. Glencoe, IL: Free Press, 1957.

Mitroff, Ian I., Richard O. Mason, and Vincent P. Barabba. *The 1980 Census: Policy Making Amid Turbulence*. Lexington, MA: D. C. Heath, 1985.

Rawls, John. *A Theory of Justice*. Cambridge, MA: Harvard University Press, 2009.

Roe, Emery. *Narrative Policy Analysis*. Durham, NC: Duke University Press, 1994.

Scriven, Michael. *Reasoning*. New York: McGraw-Hill, 1976.

Stone, Deborah. *Policy Paradox: The Art of Political Decision Making*. Rev. ed. New York: W. W. Norton, 2002.

Toulmin, Stephen Edelston. *An Examination of the Place of Reason in Ethics*. 1950.

Toulmin, Stephen Edelston. *The Uses of Argument*. Cambridge: Cambridge University Press, 2003.

Toulmin, Stephen Edelston, Robert Rieke, and Alan Janik. *An Introduction to Reasoning*. 2nd ed. New York: Macmillan, 1984.

Van Gelder, Tim. "The Rationale for Rationale." Law, Probability & Risk, 6, 1–4 (2007), pp. 23–42.

CASE 8.1 PROS AND CONS OF BALKAN INTERVENTION[59]

"Must the agony of Bosnia-Herzegovina be regarded, with whatever regrets, as somebody else's trouble? We don't think so, but the arguments on behalf of that view deserve an answer. Among them are the following:

- The Balkan conflict is a civil war and unlikely to spread beyond the borders of the former Yugoslavia. Wrong. Belgrade has missiles trained on Vienna. Tito's Yugoslavia claimed, by way of Macedonia, that northern Greece as far south as Thessaloniki belonged under its sovereignty. Those claims may return. 'Civil' war pitting non-Slavic Albanians against Serbs could spread to Albania, Turkey, Bulgaria, and Greece.

[59]From the *Los Angeles Times,* January 24, 1993, p. A7.

- The United States has no strategic interest in the Balkans. Wrong. No peace, no peace dividend. Unless the West can impose the view that ethnic purity can no longer be the basis for national sovereignty, then endless national wars will replace the Cold War. This threat has appeared in genocidal form in Bosnia. If it cannot be contained here, it will erupt elsewhere, and the Clinton administration's domestic agenda will be an early casualty.
- If the West intervenes on behalf of the Bosnians, the Russians will do so on behalf of the Serbs, and the Cold War will be reborn. Wrong. The Russians have more to fear from 'ethnic cleansing' than any people on Earth. Nothing would reassure them better than a new, post-Cold War Western policy of massive, early response against the persecution of national minorities, including the Russian minorities found in every post-Soviet republic. The Russian right may favor the Serbs, but Russian self-interest lies elsewhere.
- The Serbs also have their grievances. Wrong. They do, but their way of responding to these grievances, according to the State Department's annual human rights report, issued this past week, 'dwarfs anything seen in Europe since Nazi times.' Via the Genocide Convention, armed intervention is legal as well as justified.
- The UN peace plan is the only alternative. Wrong. Incredibly, the plan proposes the reorganization of Bosnia-Herzegovina followed by a cease-fire. A better first step would be a UN declaration that any nation or ethnic group proceeding to statehood on the principle of ethnic purity is an outlaw state and will be treated as such. As now drafted, the UN peace plan, with a map of provinces that not one party to the conflict accepts, is really a plan for continued 'ethnic cleansing.'" ■

CASE 8.2 IMAGES, ARGUMENTS, AND THE SECOND PERSIAN GULF CRISIS, 1990–1991

The analysis of policy arguments can be employed to investigate the ways that policymakers represent or structure problems (Chapter 3). We can thereby identify the images, or problem representations, that shape processes of making and justifying decisions. For example, during times of crisis, the images which United States policymakers have of another country affect deliberations about the use of peacekeeping and negotiation, the imposition of economic sanctions, or the use of deadly force. This case looks at the deliberations surrounding the U.S. decision to use military force to produce an Iraqi withdrawal from Kuwait during the Second Persian Gulf Crisis of 1990–1991.

It is important to recognize that there have been three Persian Gulf crises since 1980. The First Persian Gulf Crisis, which involved 8 years of war between Iraq and Iran, spanned the period September 1980–August 1988. The longest war in the twentieth-century, an estimated half-million civilians and military combatants died in the conflict.

The Third Gulf Crisis began in 2003 and appeared to be coming to a conclusion in 2011, as occupying U.S. and coalition troops continued a steady withdrawal from Iraq. Between 2003 and 2010, it is estimated that 700,000 to 1.5 million Iraqi civilians died as a direct and indirect result of the war, while some 11,800 members of the Iraqi security forces and police were killed.[60] There were over 4,400 U.S. military deaths and another 220 among coalition forces, about one-half of whom were British.

[60]Hannah Fischer, "Iraq Casualties: U.S. Military Forces and Iraqi Civilians, Police, and Security Forces," U.S. Congressional Research Service Report (June 11, 2010).

The Second Persian Gulf Crisis may be conveniently dated to August 2, 1990, when Iraq invaded Kuwait. Seven months later, on March 3, 1991, a cease fire was signed.[61] When Iraq invaded Kuwait in August 1990, the United States began sending troops and Patriot missiles to defend Saudi Arabia and Israel. The purpose was to defend against Iraqi SCUD missile attacks on Riyadh, Tel Aviv, and Haifa. An economic embargo was also put in place along with a U.S.–U.N. Security Council demand for an immediate withdrawal from Kuwait. Over the next 5 months Saddam Hussein failed to comply. At that point, a Security Council resolution was passed giving Iraq until January 15, 1991 to withdraw. The resolution authorized intervention by a U.S.-led coalition after that date.

Prior to the deadline for the withdrawal from Kuwait, the U.S. Senate debated the merits of United States intervention. Two policy alternatives were presented. One of these, the Dole-Warner resolution, supported the Bush administration and the United Nations by authorizing the use of military force to compel a withdrawal. The other resolution, the Mitchell-Nunn resolution, called for the continued use of sanctions. The vote in favor of the Dole-Warner resolution was 52–47.

Political scientist Richard Cottam and cognitive political psychologist James Voss show how the images of another state can be investigated by focusing on indicators of the perceived motivation, capability, culture, and decision processes of that state[62]:

- *Ally Image*: The image of the ally is known by finding in foreign policy discourse indicators referring to the mutually beneficial goals of the ally, to the ally's adequate but often less than possible military and economic capability, the ally's comparably civilized culture, and the ally's well-managed and popularly supported process of governmental decision-making.
- *Enemy Image*: The image of the enemy state is known by finding in foreign policy discourse indicators referring to the aggressive, evil, and expansionistic motivations of the enemy, to the enemy's comparable but penetrable military and economic capability, to the enemy's comparably civilized culture, and to the enemy's monolithic and undemocratic process of governmental decision-making.
- *Radical Image*: The image of the radical state is known by finding in foreign policy discourse indicators referring to fanatic or extremist motivations, to the radical state's use of terror to compensate for its inferior military and economic capability, to the radical state's less than civilized culture, and to a decision process which is well-organized and clever.
- *Degenerate Image*: The image of the degenerate state is known by finding in foreign policy discourse indicators referring to leaders motivated by the accumulation and preservation of personal power, to the degenerate state's inferior and declining military and economic capability, to the degenerate state's less than civilized culture, and to a confused and disorganized decision process.
- *Imperial*: The image of the imperial state is known by finding in foreign policy discourse indicators referring to motivations based on nationalism and modernization in the interests of the people, to the imperial state's need for external help because of inferior military and economic capability, to the imperial state's less than civilized culture, and to poorly managed decision processes which require military and economic assistance.

In the Second Persian Gulf Crisis, Senate debates were held on January 10, 11, and 12 of 1991. The arguments put forth in the debates contain indicators of these images.[63] For example, one supporter of

[61]Earlier, on February 15, the Shiite cleric Muqtada al-Sadr led a revolt against Saddam Hussein's regime, partly in response to President George H. W. Bush's earlier call on February 2 for Iraqis to overthrow Hussein's government. In the same month, the Kurds in Northern Iraq also rose up in an attempt to overthrow Saddam's regime.

[62]The images are based on Cottam, *Foreign Policy Motivation: A General Theory and a Case Study* (Pittsburgh, PA: Pittsburgh University Press, 1977). The descriptions of images and use of quotations are based on Voss et al., "Representations of the Gulf Crisis as Derived from the U.S. Senate Debate," pp. 279–302, in Donald A. Sylvan and James F. Voss, eds. *Problem Representation in Foreign Policy Decision Making* (Cambridge: Cambridge University Press, 1998).

[63]All statements were extracted from *The Congressional Record* (January 9, 10, 11) by Voss et al., "Representations of the Gulf Crisis as Derived from the U.S. Senate Debate."

the Dole-Warner resolution, which was nearly identical to the position of the White House, Senator Bryan (D-Nebraska), argued in terms indicating a traditional enemy image of Iraq: "Hussein now has under arms more men than Hitler when the German Army marched into the Rhineland . . . more tanks than when the Panzer Divisions crushed France . . . and most chilling of all, much closer to having a nuclear weapon than Adolf Hitler ever was."

This is not the image of a radical or terrorist state, but that of an aggressive, evil, and expansionistic state with capabilities comparable to the United States. Senator Roth (D-Delaware) reinforced this image: "Hussein has demonstrated that with the Cold War fading, the real threat to freedom-loving nations is the proliferation of arms in the hands of despotic dictators. Intercontinental missiles, chemical, biological, and nuclear arms can turn unstable Third World nations into first-rate military powers." Senator Cohen (R-Maine), another supporter of Dole-Warner, used an argument suggesting the image of a radical or terroristic state: "Not one of us . . . is safe from the violence currently being inflicted in Kuwait. . . . Our security is only a Pan Am 103 away at any moment." Senator Hatch (R-Utah) used similar language: "We have a major political interest in preventing Hussein from radicalizing the Arab world."

Among the supporters of the Mitchell-Nunn resolution were Senator Joseph Biden (D-Delaware), who argued: "Yes, we have interests in the Middle East. We wish to support the free flow of oil. We wish to promote stability, including the securing of Israel. But we have not heard one cogent argument that any vital American interest is at stake in a way that impels us to war." As for military capability,

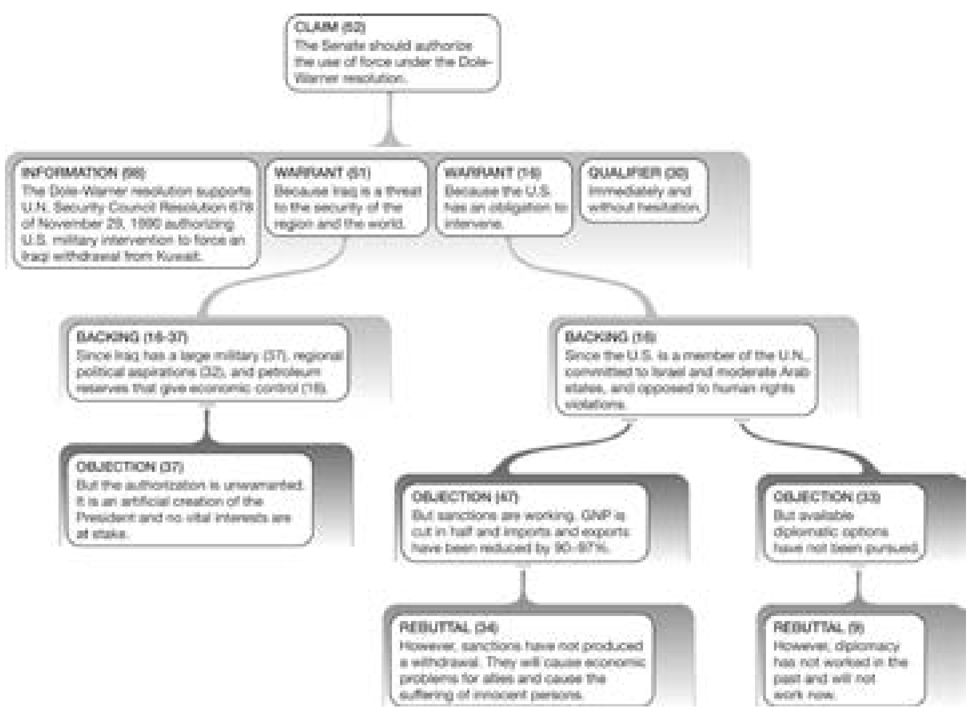


FIGURE C8.2.1
Senate Arguments Supporting and Opposing U.S. Involvement under Security Council Resolution 678 Authorizing Military Intervention to Force the Withdrawal of Iraq from Kuwait (November 29, 1990)

Senator Moynihan argued that "The Iraqis do not (even) have the technology to print their own paper money." In support of the continued use of economic sanctions, Senator Lautenberg (D-New Jersey) argued: "Historical analysis of the use of economic sanctions suggests that they can be effective over time in forcing the withdrawal of Iraqi troops from Kuwait."

Figure C8.2.1 maps the arguments supporting the resolution sponsored by Bob Dole (R-Kansas) and John Warner (R-Virginia). The figure also maps the objections of senators favoring the counter-resolution sponsored by George Mitchell (D-Maine) and Sam Nunn (D-Georgia). Rebuttals to objections are also displayed. The numbers in parentheses are the numbers of votes for the different parts of the argument supporting military intervention. The votes were along party lines, except for two Democrats who voted for the Dole-Warner Resolution. Because two senators were not present for the vote, the total number is 98 votes.

The Senate debate includes arguments that appear to be the consequence of at least three images of Iraq as a conventional enemy state, a radical or terrorist state, and a degenerate state. Arguments supporting the Dole-Warner Resolution authorizing a military attack appear to be a consequence of images of Iraq as a radical-terrorist or degenerate state. In turn, arguments offered in support of the continuation of economic sanctions appear to originate in an image of Iraq as a conventional enemy state such as Germany during World War II. The main point is that images drive problem representations, which in turn drive processes of argumentation surrounding critical issues of foreign and domestic policy. ■

CHAPTER

9

Communicating Policy Analysis

LEARNING OBJECTIVES

By studying this chapter, you should be able to:

- Describe stages in the process of policy communication
- Contrast policy analysis, documentation, communication, interactive presentation, and knowledge utilization
- Explain how the process of policy analysis is related to the process of policymaking
- Describe the main elements of policy issue papers and policy memos
- Identify factors that explain the use, misuse, and abuse of policy analysis
- Translate policy arguments into structured written narratives
- Plan, present, and evaluate an oral briefing
- Communicate a complex analysis to multiple audiences

INTRODUCTION

Policy analysis is the beginning, not the end, of efforts to improve policies. This is why policy analysis was defined in the first part of this book, not only as a process of creating and critically assessing policy-relevant information, but also of communicating this information to maximize the likelihood that it will be used. To be sure, the quality of analysis is important, but quality analysis is not necessarily used analysis. There is a large divide between the production of quality analysis and its use by policymakers (see Box 9.1).

THE PROCESS OF POLICY COMMUNICATION

The communication of policy-relevant knowledge may be viewed as a four-stage process involving analysis, documentation, communication, and utilization. As Figure 9.1 shows, policy analysis is initiated on the basis of requests for information from policymakers and other stakeholders performing different functions in the policymaking process: agenda setting, policy formulation, policy adoption, policy implementation, policy assessment, policy adaptation. These functions were discussed in Chapter 2.

In responding to requests from policymakers and other stakeholders—for example, citizens groups, nongovernmental organizations, the news media, professional bodies and interest groups—analysts create and critically assess information about policy problems, expected policy outcomes, preferred policies, observed policy outcomes, and policy performance. To do so, analysts use the various methods discussed so far under approaches to problem structuring, forecasting, prescription, monitoring, and evaluation (Chapters 3–7).

In addition to converting such information into policy arguments (Chapter 8), which are essential vehicles or carriers of policy-relevant information, analysts develop documents of different kinds: issue papers, executive summaries, memoranda, press releases, and appendices. In turn, the substance of these documents is communicated through interactions of different kinds: conferences, meetings, briefings, hearings, and electronic and social media. The purpose of developing policy-relevant documents and making presentations is to enhance prospects for the utilization of policy-relevant knowledge.

BOX 9.1 PRODUCING POLICY ANALYSIS—A POORLY MANAGED LUMBER MILL

"The social science researchers have gone into the forest of knowledge, felled a good and sturdy tree, and displayed the fruits of their good work to one another. A few enterprising, application-minded lumberjacks have dragged some logs to the river and shoved them off downstream ('diffusion' they call it). Somewhere down the river the practitioners are manning the construction companies. They manage somehow to piece together a few make-shift buildings with what they can find that has drifted down the stream, but on the whole they are sorely lacking lumber in the various sizes and forms they need to do their work properly. The problem is that someone has forgotten to build the mill to turn the logs into lumber in all its usable forms. The logs continue to pile up at one end of the system while the construction companies continue to make due at the other end. . . . There has been governmental and foundation support for the logging operation. There has also been some support for the construction companies. There has been almost nothing, however, for the planning and running of the mill." ■

Source: Rothman (1980): 16.

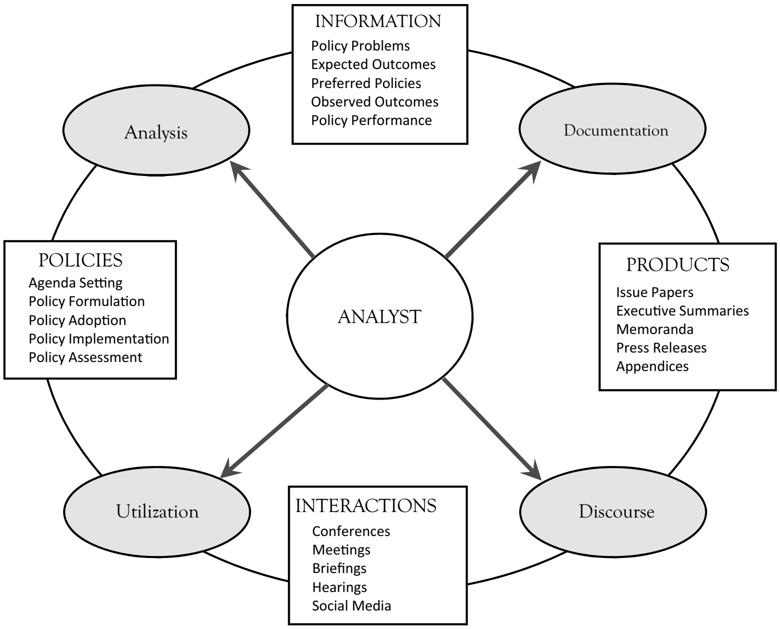


FIGURE 9.1
The Process of Policy Communication

The four arrows in Figure 9.1 indicate that analysts influence the production and use of information at multiple points, not merely by conducting analysis. Analysts not only affect the reliability, validity, and overall quality of conclusions and recommendations reached through analysis but also influence the form, content, and appropriateness of documents and subsequent interactive discourse that lead to the utilization of analysis in the policy-making process.[1] As we saw in Case 2.2 (Monitoring the Uses of Policy Analysis), there are at least four important factors that affect the use of analysis:[2]

- *Technical quality.* The reliability, validity, and appropriateness of methods is among the factors that influence the use of analysis. The movement towards evidence-based policy affirms the importance of technical quality, notwithstanding continuing debates about the definition of evidence and science itself.

[1]The recent National Academy of Sciences Report, *Using Science as Evidence in Public Policy* (Washington, DC: National Academies Press, 2012), is an informative source of illustrations of this debate. The report was edited by Kenneth Prewitt, Thomas A. Schwandt, and Miron L. Straf on behalf of the Committee on the Use of Social Science Knowledge in Public Policy, Center for Education, Division of Behavioral and Social Sciences and Education, National Research Council.

[2]Carol H. Weiss and Michael J. Bucuvalas. "Truth Tests and Utility Tests: Decision-Makers' Frames of Reference for Social Science Research," *American Sociological Review* (1980): 302–313. For a review and critical assessment of literature on this issue, see William N. Dunn and Burkart Holzner, "Knowledge in Society: Anatomy of an Emergent Field," *Knowledge in Society* 1, 1 (1988): 3–26. See also William N. Dunn, "The Two-Communities Metaphor and Models of Knowledge Use: An Exploratory Case Survey," *Science Communication* 1, 4 (1980): 515–536. David J. Webber, "The Distribution and Use of Policy Knowledge in the Policy Process," *Knowledge, Technology & Policy* 4, 4 (1991): 6–35. Judith A. Bradbury, "The Policy Implications of Differing Concepts of Risk," *Science, Technology & Human Values* 14, 4 (1989): 380–399.

- *Conformity to user expectations.* Users of analysis do not have uniform expectations about the usefulness of analysis, nor do they use the same criteria for assessing the truth and utility of information.
- *Challenge to the status quo.* Despite the absence of uniform expectations, users appear to be influenced by research-based information that challenges existing policies. Challenges to existing policies may be consistent with varied expectations about the truth or utility of information.
- *Organizational constraints.* The several models of policy change and stability discussed in Chapter 2, including bounded rationality and partisan mutual adjustment, affirm that social, political, and economic factors external to the process of assessing the quality, truth, or utility of information play an important role in shaping the use of information.
- *Actionability.* A key feature of policy analysis and other branches of the applied social sciences is that they are unlikely to be acted on unless they are useful in some aspect of policymaking. The foundations of policy analysis, as represented by the policy sciences, are dependent on a view of knowledge that originated in the theories of Dewey and other pragmatists.

Tasks in Policy Documentation

The knowledge and skills needed to conduct policy analysis are different from those needed to develop policy-relevant documents. The development of policy-relevant documents capable of conveying usable knowledge requires knowledge and skills in synthesizing, organizing, translating, simplifying, displaying, and summarizing information.

Synthesis. Analysts typically work with hundreds of pages of previously published reports, newspaper and journal articles, summaries of interviews with stakeholders and other key informants, copies of existing and model legislation, tables of statistical series, and more. Information must be synthesized into documents ranging in length from one or two pages (policy memoranda) to more than twenty-five pages in length (policy issue papers and reports). Information also must be synthesized when summarizing policy issue papers in the form of executive summaries and news releases.

Organization. Analysts must be able to organize information in a coherent, logically consistent, and economical manner. Although documents vary in style, content, and length, they typically have certain common elements:

- Overview or summary of contents of paper
- Background of previous efforts to solve the problem
- Diagnosis of the scope, severity, and causes of the problem
- Identification and evaluation of alternative solutions
- Recommendations of actions that may solve the problem
- Policy issue papers, as contrasted with policy memos, usually including additional elements with tables and graphs, technical appendices that explain results of data analysis, existing and proposed legislation, descriptions of formulas and equations, and other supporting material

Translation. The specialized terminology and techniques of policy analysis must be translated into the languages of policy stakeholders. In many cases, this requires the conversion of abstract theoretical concepts and complex analytical and statistical routines into the ordinary language and arguments of non-experts. When the audience includes experts on the problem (e.g., other analysts and staff specialists), a detailed exposition of theoretical concepts and analytical and statistical routines can be incorporated in appendices.

Simplification. Potential solutions to a problem are often complex. The combinations and permutations of policy alternatives and outcomes can easily exceed the hundreds. In such cases, a smaller set of options and outcomes can be displayed in the form of a matrix or "scorecard."[3] Another way to simplify complexity is to use a strategy table as part of a decision tree or to use a decision tree. Both scorecards and decision trees were presented in Chapter 1. The simplification of complex quantitative relationships also can be accomplished by presenting in ordinary language cases that typify quantitative profiles.[4]

Visual displays. The availability of advanced, user-friendly computer graphics has dramatically increased capabilities of effective visual communication. The visual display of quantitative information—bar charts, histograms, pie charts, line graphs, maps generated from geographic information systems (GIS)—is essential for effective policy communication.[5]

Summaries. Policymakers with crowded agendas operate under severe time constraints that limit their reading to no more than a few minutes a day. Members of the United States Congress spend approximately 15 minutes per day reading, and most of this time is spent on local and national newspapers.[6] The same is true for higher-level policymakers in agencies such as the U.S. Department of State. Under these circumstances, staff analysts perform an essential role in briefing policymakers, who are far more likely to read an executive summary or condensed memorandum than a full policy issue paper. Skills in preparing summaries are essential for effective policy communication. The most comprehensive and detailed document is the policy issue paper. A policy issue paper addresses questions such as the following:

- In what ways can the policy problem be formulated?
- What is the scope and severity of the problem?
- To what extent does it require public action?
- If no action is taken, how is the problem likely to change in coming months or years?

[3]The scorecard or matrix displays developed by Bruce F. Goeller of the RAND Corporation are useful as means for simplifying a large set of interdependent alternatives. For an overview of the Goeller scorecard or matrix, see Bruce F. Goeller, "A Framework for Evaluating Success in Systems Analysis" (Santa Monica, CA: RAND Corporation, 1988).

[4]A good example is Ronald D. Brunner, "Case-Wise Policy Information Systems: Redefining Poverty," *Policy Sciences,* 19 (1986): 201–223.

[5]The outstanding source on the methodology of graphic displays is Edward R. Tufte's trilogy, *The Visual Display of Quantitative Information* (Cheshire, CT: Graphics Press, 1983); *Envisioning Information* (Cheshire, CT: Graphics Press, 1990); *Visual Explanations* (Cheshire, CT: Graphics Press, 1997).

[6]Personal communications with staff of the U.S. General Accounting Office.

- Have other units of government addressed the problem, and if so, what were the consequences?
- What goals and objectives should be pursued in solving the problem?
- What major policy alternatives are available to achieve these goals and objectives?
- What criteria should be employed to evaluate the performance of these alternatives?
- What alternative(s) should be adopted and implemented?
- What agency should have the responsibility for policy implementation?
- How will the policy be monitored and evaluated?

Analysts are seldom requested to provide answers to all these questions. Instead, they are asked to address a smaller set of these questions that have arisen in one or several phases of the policymaking process—for example, in the agenda-setting phase, questions about the future costs, benefits, and availability of health care may arise. Policy issue papers, it should be noted, are less frequently requested than short policy memoranda or policy briefs ranging from one to several pages in length. Policy memoranda and briefs draw on and synthesize the substance, conclusions, and recommendations of policy issue papers, research reports, and other source documents. News releases, in turn, usually summarize the conclusions and recommendations of a major policy issue paper or report.

The multiplicity of policy documents draws attention to the fact that there are many ways to develop appropriate written materials on the basis of the same policy analysis. There are also multiple audiences for policy-relevant information. Immediate "clients" are often only one audience, and effective communication may demand that analysts develop different documents for different audiences, thus thinking strategically about opportunities for policy improvement. "Thinking strategically about the composition of the audience is essential for effective communication. The selection of the audience is not something to be left solely to the immediate client. . . . There are many clients to be reached; some may be peripheral, some remote, some in the future, yet all are part of a potential audience."[7] For example, when the preparation of news releases is permitted under standard operating procedures, the news release is the most appropriate vehicle for reaching the general public through the mass media.[8] A policy issue paper is not. If the aim is to communicate with the immediate client, however, it is the executive summary or policy memorandum that is likely to be most effective.

Tasks in Oral Presentations and Briefings

Just as procedures for conducting analysis are different from procedures for developing policy-relevant documents, so are procedures for developing these documents different from procedures for their communication in the form of oral presentations and briefings.

[7]See Arnold Meltsner, "Don't Slight Communication," *Policy Analysis* 2, 2 (1978): 221–231.

[8]On the role of the mass media in communicating social science knowledge, see Carol H. Weiss and Eleanor Singer, with the assistance of Phyllis Endreny, *Reporting of Social Science in the National Media* (New York: Russell Sage Foundation, 1987). A much-neglected area, the role of academic and commercial publishing in communicating ideas emanating from the sciences and humanities, is addressed in Irving Louis Horowitz, *Communicating Ideas: The Crisis of Publishing in a Post-Industrial Society* (New York: Oxford University Press, 1986).

A common medium of communication is the mailed document, an impersonal means of reaching clients and other policy stakeholders by physically transmitting the document. The major limitation of this medium is the probability that a document will reach intended beneficiaries—but then sit on the shelf. The probability of utilization is enhanced when ideas are communicated through policy presentations. Policy presentations, which include conversations, conferences, briefings, meetings, and hearings, constitute an interactive mode of communication that is positively associated with the utilization of policy-relevant knowledge.[9]

Although there is no codified body of rules for making oral presentations, experience has shown that a number of general guidelines are important for effective policy communication. These guidelines offer multiple communications strategies appropriate for the various contingencies found in complex practice settings. Among these contingencies are:

- The size of the group that comprises the audience
- The number of specialists in the problem area addressed
- The familiarity of group members with methods of analysis
- The credibility of the analyst to the group
- The extent to which the presentation is critical to policies under active consideration

BOX 9.2 CONTINGENT COMMUNICATION

Policy analyses are frequently presented to groups that have few experts in the problem area, minimum familiarity with analytic methods, limited confidence in analysts, and little time for meetings that take up precious time. What communications strategies are likely to be effective under these conditions?*

- Make sure that the presentation addresses the needs of key decision-makers and recognizes audience diversity.
- Avoid giving too much background information.
- Focus on conclusions.
- Use simple graphics to display data.
- Discuss methods only if necessary to support conclusions.
- Pinpoint reasons for your lack of credibility, choosing a strategy to overcome the problem—for example, arrange to be introduced by a credible associate or present as part of a team.
- Be sensitive to time constraints and the possibility that the group is already committed to a course of action.
- Position your supporters next to people with anticipated negative reactions.
- Prioritize your points so that you present those that are most critical to the group's preferred decision.

Note: *Adapted from Version 2.0 of *Presentation Planner,* a software package developed by Eastman Technology, Inc.

[9]The efficacy of the "interactive" model of communicating and using information has been reported in literature for at least 30 years. See Ronald G. Havelock, *Planning for Innovation: Through Dissemination and Utilization of Knowledge* (Ann Arbor, MI: Institute for Social Research, Center for the Utilization of Scientific Knowledge, 1969); Carol H. Weiss, "Introduction," in Using *Social Research in Public Policy Making* (Lexington, MA: D. C. Heath, 1977), pp. 1–22; Charles E. Lindblom and David Cohen, *Usable Knowledge: Social Science and Social Problem Solving* (New Haven, CT: Yale University Press, 1979); and Michael Huberman, "Steps toward an Integrated Model of Research Utilization," *Knowledge: Creation, Diffusion, Utilization* 8, 4 (June 1987): 586–611.

In such contexts, multiple communications strategies are essential: There is no "universal client" who uses the same standards to evaluate the plausibility, relevance, and usability of policy analysis. Effective policy presentations are contingent on matching communications strategies to characteristics of the audience for policy analysis (Box 9.2).

THE POLICY ISSUE PAPER

A policy issue paper should provide answers to a number of questions:

- What actual or potential courses of action are the objects of conflict or disagreement among stakeholders?
- In what different ways may the problem be defined?
- What is the scope and severity of the problem?
- How is the problem likely to change in future?
- What goals and objectives should be pursued to solve the problem?
- How can the degree of success in achieving objectives be measured?
- What activities are now under way to resolve the problem?
- What new or adapted policy alternatives should be considered as ways to resolve the problem?
- Which alternative(s) are preferable, given the goals and objectives?

In answering these questions, analysts need to acquire knowledge and skills in writing policy issue papers. Only recently have such knowledge and skills become integral parts of instructional programs in university departments and professional schools of public policy, public management, and public administration. Policy analysis, although it draws from and builds on social science disciplines, differs from those disciplines because it seeks to improve as well as understand policymaking processes. In this respect, policy analysis has characteristics that make it an "applied" rather than "basic" policy discipline.[10] Analysts need to acquire knowledge and skills provided in basic policy disciplines. But the purpose of policy analysis remains "applied" rather than "basic." Characteristics of basic and applied policy disciplines are displayed in Table 9.1.

Issues Addressed in the Policy Issue Paper

The policy issue paper may address issues in almost any area: health, education, welfare, crime, labor, energy, foreign aid, national security, human rights, and so on. Papers in any one of these issue areas may focus on problems at one or more levels of government. Air pollution, global warming, and terrorism, for example, are international, national, and local in scope. Issue papers may be presented as "staff reports," "briefing papers," "options

[10]The terms "basic" and "applied" are used for convenience only. It is widely acknowledged that these two orientations toward science overlap in practice. Many of the most important advances in the "basic" social sciences originated in "applied" work on practical problems and conflicts. See, especially, Karl W. Deutsch, Andrei S. Markovits, and John Platt, *Advances in the Social Sciences, 1900–1980: What, Who, Where, How?* (Lanham, MD: University Press of America and Abt Books, 1986).

TABLE 9.1

Two Kinds of Policy Disciplines

Characteristic	Basic	Applied
Origin of problems	University colleagues	Clients and citizens
Typical methods	Quantitative modeling	Develop arguments
Type of research	Original data collection	Synthesis of existing data
Primary aim	Improve theory	Improve practice
Communications media	Article or book	Policy memo or issue paper
Source of incentives	Universities	Governments and NGOs

papers," or so-called "white papers." An illustrative list of issues that may be the focus of a policy issue paper is presented below.

- Which of several contracts should a union bargaining team accept?
- Should the mayor increase expenditures on road maintenance?
- Should the city manager install a computerized management information system?
- Which public transportation plan should the mayor submit for federal funding?
- Should a state agency establish a special office to recruit minorities and women for civil service positions?
- Should a citizens' group support environmental protection legislation now before Congress?
- Should the governor veto a tax bill passed by the state legislature?
- Should an agency director support a plan for flexible working hours (flextime)?
- Should a legislator support a bill restricting the sale of handguns?
- Should the president withhold foreign aid from countries that violate human rights?
- Should the United Nations General Assembly condemn the violation of human rights in a particular country?
- Should the United States withdraw from the International Labour Organization?
- Should taxes on foreign investments of multinational corporations registered in the United States be increased?

Elements of the Policy Issue Paper

A policy issue paper should "explore the problem at a depth sufficient to give the reader a good idea of its dimensions and the possible scope of the solution, so that it might be possible for a decision-maker to conclude either to do nothing further or to commission a definitive study looking toward some action recommendation."[11] In this author's experience, most issue papers deal primarily with the formulation of a problem and possible solutions. Rarely does the issue paper reach definitive conclusions or recommendations. While an issue paper may contain recommendations and outline plans for monitoring and

[11]E. S. Quade, *Analysis* for *Public Decisions* (New York: American Elsevier Publishing, 1975), p. 69.

evaluating policy outcomes, it is essentially the first phase of an in-depth policy analysis that may be undertaken at a later time.

In preparing an issue paper, the analyst should be reasonably sure that all major questions have been addressed. Although issue papers vary with the nature of the problem being investigated, most issue papers contain a number of standard elements.[12] These elements have been organized around the framework for policy analysis presented in the text.

Observe that each element of the issue paper requires different policy analytic methods to produce and transform information. A policy issue paper, however, is essentially a prospective (*ex ante*) investigation. It is based on limited information about past actions, outcomes, and performance and, in this respect, differs from program evaluations and other retrospective (*ex post*) studies. Appendix 1 provides a checklist for preparing a policy issue paper.

The Policy Memorandum

It is widely but mistakenly believed that policy analysts spend most of their time developing policy issue papers, studies, and reports. In fact, the primary activity of most analysts is the preparation of policy memoranda. Whereas the issue paper is a long-term activity involving the conduct of policy research and analysis over many months, the policy memorandum is prepared over a short period of time—usually no more than 1 month, but often a matter of days. The differences between the issue paper, study, or report and the policy memorandum reflect some of the differences between basic and applied policy analysis summarized in Table 9.1.

The policy memorandum should be concise, focused, and well organized. It reports background information and presents conclusions or recommendations, usually in response to a request from a client. The policy memo often summarizes and evaluates one or more issue papers or reports, along with accompanying backup documents and statistical data.

The format for a memo, which varies from agency to agency, is designed for quick and efficient reading. Most memos display the name of the recipient, the name of the analyst submitting the memo, the date, and the subject on separate lines at the top of the page. Most agencies have preprinted forms for memos. Word processing programs have templates for memos and other standard documents. Writers of policy memo should take every opportunity to communicate clearly and effectively the main points of the analysis:

- The subject line should state in concise terms the main conclusion, recommendation, or purpose of the memo.
- The body of the memo (usually no more than two pages) should contain headings.
- Bulleted lists should be used to highlight a small set (no more than five) of important items such as goals, objectives, or options.
- The introductory paragraph should review the request for information and analysis, restate the major question(s) asked by the client, and describe the objectives of the memo.

A sample policy memorandum is presented in Appendix 3.

[12]Compare Eugene Bardach, *The Eight-Step Path to Policy Analysis: A Handbook for Practice* (Berkeley: University of California Press, 1996); Quade, *Analysis for Public Decisions*, pp. 68–82; and Harry Hatry and others, *Program Analysis for State and Local Governments* (Washington, DC: Urban Institute, 1976), appendix B, pp. 139–143.

The Executive Summary

The executive summary is a synopsis of the major elements of a policy issue paper or report. The executive summary typically has the following elements:

- Purpose of the issue paper or study being summarized
- Background of the problem or question addressed
- Major findings or conclusions
- Approach to analysis and methodology (where appropriate)
- Conclusions and recommendations (when requested)

Appendix 2 provides a sample executive summary.

The Letter of Transmittal

A letter of transmittal will accompany a policy issue paper or study. Although the letter of transmittal has most of the elements of a policy memo, the purpose of the letter (or memo) is to introduce the recipient to a larger paper or study. Some common elements of a letter of transmittal are as follows:

- The letterhead or address of the person(s) transmitting the issue paper or study
- The name, formal title, and address of the person(s) or client(s) who requested the issue paper or study

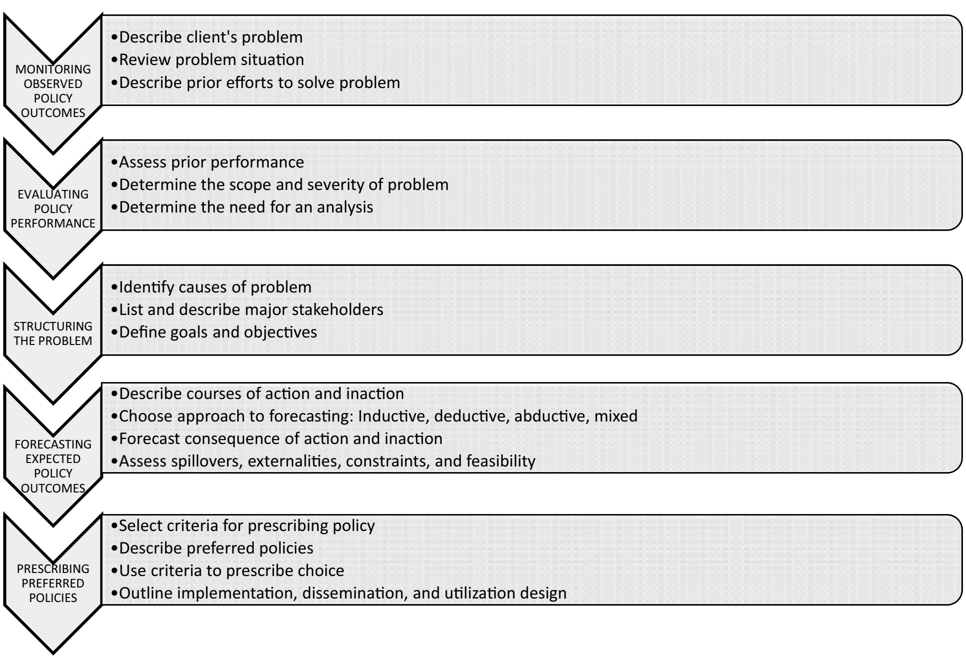


FIGURE 9.2
Methods for Developing a Policy Memo or Issue Paper

- A short paragraph stating the question or problem that the person(s) or client(s) expect the analyst to address
- A brief summary of the most important conclusions or recommendations of the issue paper or study
- A review of arrangements for further communication or work with the client, for example, a scheduled briefing on the issue paper or study
- A concluding statement indicating where and how the analyst can be reached to answer any questions
- A signature with the name and title of the analyst or the analyst's supervisor.

PLANNING THE ORAL BRIEFING AND VISUAL DISPLAYS

Written documents are only one means for communicating the results of policy analysis. The other is the oral briefing. Although the oral briefing or presentation has many of the same elements as the issue paper, the approach to planning and delivering an oral presentation is quite different. The elements of an oral briefing typically include the following:

- Opening and greeting to participants
- Background of the briefing
- Major findings of issue paper, study, or report
- Approach and methods
- Data used as basis for the analysis
- Recommendations
- Questions from participants
- Closing

Knowing the audience is one of the most important aspects of planning oral briefings. In this context, a series of important questions should be answered prior to the briefing.

- How large is the audience?
- How many members of the audience are experts on the problem you are addressing in the briefing?
- What percent of the audience understands your research methods?
- Are you credible to members of the audience?
- Does the audience prefer detailed or general information?
- Where does the briefing fall in the policymaking process? At the agenda-setting phase? The implementation phase?

Answers to these questions govern the strategy and tactics of communication appropriate for the particular audience. For example, if the audience is a medium-size group with twenty to twenty-five persons, the diversity of the audience calls for specific communications strategies and tactics:

- Circulate background materials before the meeting so that everyone will start from a common base.

- Tell the group that the strategy selected is designed to meet the objectives of the group as a whole but that it may not satisfy everyone's needs.
- Focus on the agendas of key policymakers, even if you lose the rest of the audience.

Another common problem in giving oral briefings is that policymakers may bring their own staff experts to the briefing. If a high percentage of the audience has expertise in the area you are addressing, the following strategies and tactics are appropriate.

- Avoid engaging in a "knowledge contest" with subject experts. Focus instead on the purpose of the presentation.
- Capitalize on the group's expertise by focusing on findings and recommendations.
- Avoid a lengthy presentation of background material and description of methods.
- Attempt to generate dialogue and a productive debate on policy options.

An equally important problem in oral briefings is knowing individual members of the group. For example, it is important to know if the immediate clients for policy analysis are:

- Experts in the problem area
- Familiar with your analytic methods
- Influential participants in the policy process
- Oriented toward detailed versus broad information
- Have a high, medium, or low stake in the outcome
- Politically important to the success of your recommendations

Finally, the effectiveness of oral briefings can be significantly enhanced by various graphic displays. The most common of these are transparencies used with overhead projectors, and slides created by computer graphics and presentation programs such as Microsoft PowerPoint. In using overhead transparencies and slides, it is important to recognize that they are not appropriate when you want an informal style of presentation—overheads tend to make briefings very formal and make open dialogue difficult. It is also important to observe the following guidelines:

- Keep transparencies and slides simple, with condensed and highlighted text.
- Use clear, bold, uncluttered letters and lines of text.
- Avoid complicated color schemes, pop art, and fancy templates.
- Highlight important points and use different colors. A dark background with superimposed white or yellow letters is effective.
- Limit the text on any page to a maximum of ten lines.
- Be sure the width of the screen is at least one-sixth of the distance between the screen and the farthest viewer.
- Make sure that the nearest viewer is a minimum of two screen widths from the screen.
- Remain on the same transparency or slide for a minimum of 2 minutes. This gives viewers time to read, plus an additional 30 seconds to study the content.
- Leave the room lights *on*—transparencies and slides are designed for lighted rooms. It is not a movie theater.

- Turn off the overhead projector, or blacken the computer screen, when you want to redirect attention back to you and reestablish eye contact.

These are but a few of the strategies and tactics that are appropriate for different audiences and communications media. Given the goal of selecting communications media and products that match the characteristics of the audience, there is today a rich array of techniques for the visual display of information. Appropriate computer graphics programs are widely available.

Many types of visual displays are available to the analyst who wishes to effectively communicate complex ideas. Examples of the following visual displays (created with *Harvard Graphics*) are provided in Appendix 4. The same graphics can be created with Microsoft Excel and PowerPoint.

- Options-impact matrices ("Goeller scorecards")
- Spreadsheets
- Bar charts
- Pie charts
- Frequency histograms
- Ogives (cumulative frequency curves)
- Scatterplots
- Interrupted time-series and control-series graphs
- Influence diagrams
- Decision trees
- Strategy tables
- Causal (arrow) diagrams

POLICY ANALYSIS IN THE POLICYMAKING PROCESS

We began this book by defining policy analysis as a process of multidisciplinary inquiry designed to create, critically assess, and communicate information that is useful in understanding and improving policies. Throughout, we have emphasized that policy analysis is a series of intellectual activities embedded in a social process known as policymaking. The distinction between these two dimensions—the intellectual and the social—is important for understanding the use, underuse, and nonuse of analysis by policymakers.

Recall the three interrelated dimensions of information use presented in Chapter 2:

- *Composition of users.* Analysis is used by individuals as well as collectives (e.g., agencies, bureaus, legislatures). When the use of analysis involves individual gains (or losses) in the perceived future value of information, the process of use is an aspect of individual decisions (*individual use*). By contrast, when the process of use involves public enlightenment or collective learning, the use of information is an aspect of collective decisions (*collective use*).
- *Effects of use.* The use of policy analysis has different kinds of effects. Policy analysis is used to think about problems and solutions (conceptual use). It is also used to legitimize preferred formulations of problems and solutions by invoking the

authority of experts, religion, and methods (symbolic use). By contrast, behavioral effects involve the use of policy analysis as a means or instrument for carrying out observable policymaking activities (instrumental use). Conceptual, symbolic, and behavioral uses occur at individual and collective levels.

- *Scope of information used.* The scope of use ranges from the specific to the general. The use of "ideas in good currency" is general in scope (general use), whereas the use of a policy recommendation is specific (specific use). Information that varies in scope is used by individuals and collectives, with effects that are conceptual, symbolic, and behavioral.

With these distinctions in mind, the use of information in policymaking is shaped by at least five types of factors.[13] These factors include differences in the characteristics of information, the modes of inquiry use to produce information, the structure of policy problems, political and bureaucratic structures, and the nature of interactions among policy analysts, policymakers, and other stakeholders.

Perceived Quality of Information

The characteristics of information produced by policy analyses influences its use by policymakers. Information that conforms to the product specifications of policymakers is more likely to be used than information that does not, because product specifications reflect policymaker needs, values, and perceived opportunities. Policymakers tend to value information that is communicated in personal verbal reports, rather than formal written documents, and expressed in a language that reflects concrete policy contexts rather than the more abstract vocabularies of the natural and social sciences.[14] Policymakers also attach greater value to information that is accurate, precise, and generalizable to similar settings.[15]

Perceived Quality of Methods

The use of information by policymakers is shaped by the process of inquiry used by analysts to produce and interpret that information. Information that conforms to standards of quality research and analysis is more likely to be used. Yet there are widely differing views about the meaning of "quality." For many analysts, quality is defined in terms of the use of social experimentation, random sampling, and quantitative measurement.[16] The assumption is that information use is a function of the degree to which policy research and analysis conforms to accepted scientific methods, provided that the resultant information is adapted to organizational constraints such as the need for timely information.

[13]For case study research on these factors, see William N. Dunn. "The Two-Communities Metaphor and Models of Knowledge Use: An Exploratory Case Survey," *Knowledge: Creation, Diffusion, Utilization*, 4 (June 1980).
[14]See Mark van de Vall, Cheryl Bolas, and Thomas Kang, "Applied Social Research in Industrial Organizations: An Evaluation of Functions, Theory, and Methods," *Journal of Applied Behavioral Science* 12, 2 (1976): 158–177.
[15]See Nathan Caplan, Andrea Morrison, and Roger J. Stambaugh, *The Use of Social Science Knowledge in Policy Decisions at the National Level* (Ann Arbor, MI: Institute for Social Research, 1975).
[16]See I. N. Bernstein and H. E. Freeman, *Academic and Entrepreneurial Research* (New York: Russell Sage Foundation, 1975).

By contrast, other analysts define quality in different terms. Here, quality is defined in terms that emphasize nonquantitative procedures designed to uncover subjective understandings about problems and potential solutions held by policymakers and other stakeholders.[17]

Complexity of Problems

The utilization of information by policymakers is also influenced by the goodness-of-fit between methods and the complexity of problems. Relatively well-structured problems involving consensus on goals, objectives, alternatives, and their consequences require different methodologies than relatively ill-structured problems. Because the essential characteristic of ill-structured problems is conflict, not consensus, ill-structured problems require methodologies that are holistic and that bring to bear multiple perspectives of the same problem situation in defining the nature of the problem itself.[18]

The distinction between well-structured and ill-structured problems is associated with the distinction between lower-level (micro) and higher-level (meta) problems. Meta-level problems involve issues of how to structure a problem, how to assist policymakers in knowing what should be known, and how to decide which aspects of a problem should be resolved on the basis of "conceptual" knowledge.[19] Most policy research and analysis is oriented toward the production of "instrumental" knowledge, that is, knowledge about the most appropriate means for attaining taken-for-granted or agreed-upon ends. Instrumental knowledge, since it is appropriate for well-structured problems at the micro level, is less likely to be used by policymakers who face ill-structured problems involving disagreements about the nature of problems and their potential solutions.

Political and Organizational Constraints

The use of information is also shaped by differences in the formal structures, procedures, and incentive systems of public organizations. The influence of policymaking elites, the bureaucratization of roles, the formalization of procedures, and the operation of incentive systems that reward conservatism and punish innovation contribute to the underuse and nonuse of information produced by analysts. These and other factors, although they are "external" to the methodology of policy analysis, create a political and bureaucratic context for information use.[20]

[17] Martin Rein and Sheldon H. White, "Policy Research: Belief and Doubt," *Policy Analysis* 3, 2 (1977); Michael Q. Patton, *Alternative Evaluation Research Paradigm* (Grand Forks: University of North Dakota, 1975); and H. Aeland, "Are Randomized Experiments the Cadillacs of Design?" *Policy Analysis* 5, 2 (1979): 223–242.

[18] Russell Ackoff, *Redesigning the Future: A Systems Approach to Societal Problems* (New York: Wiley, 1974); Ian I. Mitroff, *The Subjective Side of Science* (New York: American Elsevier Publishing, 1974); and Ian I Mitroff and L. Vaughan Blankenship, "On the Methodology of the Holistic Experiment: An Approach to the Conceptualization of Large-Scale Social Experiments," *Technological Forecasting and Social Change* 4 (1973): 339–353.

[19] See Nathan Caplan, "The Two-Communities Theory and Knowledge Utilization," *American Behavioral Scientist* 22, 3 (1979): 459–470.

[20] Robert F. Rich, *The Power of Social Science Information and Public Policymaking: The Case of the Continuous National Survey* (San Francisco, CA: Jossey-Bass, 1981).

Interactions among Policymakers and Other Stakeholders

The nature and types of interaction among stakeholders in the various phases of policymaking also influence the use of information by policymakers.[21] Policy analysis is not simply a scientific and technical process; it is also a social process where the structure, scope, and intensity of interaction among stakeholders govern the creation and use of information.

The interactive nature of policy analysis makes questions of information use complex. Analysts rarely produce information that is or can be used merely for "problem solving," in the narrow sense of discovering through analysis the most appropriate means to ends that are well defined and about which there is substantial consensus. Many of the most important policy problems, as we have seen in this book, are sufficiently ill structured that the "problem-solving" model of policy analysis is inappropriate or inapplicable. For this reason, policy analysis has been described throughout this book as an integrated process of inquiry where multiple methods—problem structuring, forecasting, prescription, monitoring, evaluation—are used continuously to create and transform information. Although the process of policy analysis is quintessentially methodological, it is also a process of communicative interaction. Policy analysis, because it aims at the creation, critical assessment, and communication of policy-relevant information, is vital to policy argumentation and public debate.

CHAPTER SUMMARY

This concluding chapter has provided an overview of the process of policy communication and its importance to the use of analysis by policymakers. Policy analysis is the beginning, not the end, of efforts to improve policymaking. Knowledge and skills in conducting policy analysis are different from knowledge and skills in developing policy documents and giving oral briefings. To be effective, analysts need to master and apply a broad range of communications skills, thereby narrowing the great divide between intellectual and social dimensions of policymaking.

REVIEW QUESTIONS

1. Why are the communication and use of information central to the aims of policy analysis?
2. If policy analysis is an intellectual activity carried out within a social process, what are the implications for the use of analysis by policymakers?
3. "Policy analysis is the beginning, not the end, of efforts to improve the policymaking process." Comment.
4. Before intended beneficiaries can use policy-relevant information, it must be converted into policy-relevant documents and communicated in presentations of different kinds. Does this guarantee that intended beneficiaries will use information?
5. Describe stages of the process of policy communication.

[21]See Weiss, "Introduction," *Using Social Research in Public Policy Making*, pp. 13–15.

6. Why are skills needed to develop policy documents and give oral presentations different from skills needed to conduct policy analysis?
7. Discuss factors that affect the use of policy analysis by policymakers and other stakeholders.
8. Why is the production of policy analysis described as a poorly managed lumber mill?

DEMONSTRATION EXERCISES

1. After reading Case 9.1, create argument maps for the four types of documents described in the case:
   - A position paper
   - An issue paper or memo
   - A policy brief
   - A letter to the editor

   After creating the maps, develop written documents that include the reasoning used in each of the four maps.
2. The purpose of this exercise is to sharpen knowledge and skills in planning and conducting effective oral briefings. This is a group exercise, although different individuals could also complete it. Divide the class into four groups. Each group should perform the following tasks.
   a. Respond to one of the following requests (I, II, III, or IV) to give an oral briefing on the reduction of harmful levels of lead in gasoline. Assume that you are the authors of the analysis included as Case 9.2. It is an analysis prepared by David Weimer and Aidan Vining, two highly experienced policy analysts. Your group has been asked to make one of the four briefings below.
   b. Each group will have fifteen minutes to give its briefing and respond to ten minutes of questioning. The main policy question you must answer is: Should the government regulate lead additives in gasoline?
   c. Other class members, who will be using the attached rating scale, will evaluate presentations. The evaluations will be used as a basis for a critique and discussion of the briefings.
   - *Group I.* Prepare an oral briefing for a community environmental group composed of sixty citizens who have little or no knowledge of statistical methods and biomedical research on lead. The audience does not appear to your group, which is composed of "outsiders" from Washington. Among members of the group is a potentially disruptive person who likes to get attention by engaging in a "knowledge contest." This person is running for public office in the next election. The environmental group wants to see new legislation on the public agenda. Your answer to the policy question stated earlier is supposed to help them.
   - *Group II.* Prepare an oral briefing for a group of ten experts in biostatistics, econometrics, and environmental economics. All have extensive knowledge of the health effects of leaded gasoline. Most of them are colleagues with whom you have worked previously. One member of the working group, a respected biostatistician, is a paid consultant working for the International Lead and Zinc Union (ILZU), an organization that markets lead and zinc products. The ILZU takes the position that "statistical correlations do not prove causation." The working group is designing a research proposal to investigate the health effects of lead. Your answer to the policy question is supposed to help them.
   - *Group III.* Prepare an oral briefing for a group of thirty business executives from the oil refining industry. They are opposed to government regulation of leaded gasoline. They believe that regulation will reduce profits and destroy the industry. Although they have been briefed by their staff experts, they are not familiar with statistics and biomedical research on lead. The executives

are preparing to fight a congressional bill that, if adopted next month, would regulate and eventually ban leaded gasoline. Your answer to the policy question is supposed to help them.

- *Group IV.* Prepare an oral briefing for a group of six experts from the Occupational Safety and Health Administration. They have done extensive research on the biomedical effects of lead exposure, and are convinced that low levels of exposure to lead are extremely harmful, especially to inner city children who are exposed to unusually high amounts of automotive exhaust. The experts are evaluating your analysis in preparation for an annual conference. Your answer to the policy question is supposed to help them.

In preparing your briefing, use the structural model of argument (Chapter 8) as a framework for organizing your presentation, clearly stating your argument, and anticipating challenges. (*Note:* Do not present the model of argument as part of the briefing. It is an analytic tool.)

Your oral briefing should be based on a communications strategy appropriate to the characteristics of your audience. To develop an appropriate communications strategy, use the guidelines described in this chapter. The briefings will be evaluated on the basis of the attached checklist and rating scales. You or your instructor should make four (4) copies of the checklist—three to evaluate the other presentations and one for your records.

BOX 9.3 CHECKLIST FOR AN ORAL BRIEFING

Group evaluated (circle): I II III IV

Use the following rating scale for your evaluation. Provide a brief written justification in the space provided next to each item.

1 = Very good
2 = Good
3 = Adequate
4 = Poor
5 = Very poor

1. Rate the effectiveness of the following elements of the briefing:
   (a) opening __________
   (b) background __________
   (c) findings __________
   (d) methods __________
   (e) evidence/data __________
   (f) recommendations __________
   (g) questions and answers __________
   (h) closing __________
2. Appropriateness of briefing to:
   (a) size of the audience __________
   (b) expertise of audience __________
   (c) familiarity of audience with methods __________

(*continued*)

(d) audience information preferences __________
(e) phase of policymaking process __________
(f) perceptions of presenter's credibility __________

3. Logic, organization, and flow of briefing: __________
4. Use of visual displays: __________
5. Overall, this oral briefing was: __________

BIBLIOGRAPHY

Barber, Bernard. *Effective Social Science: Eight Cases in Economics, Political Science, and Sociology.* New York: Russell Sage Foundation, 1987.

Freidson, Elliot. *Professional Powers: A Study of the Institutionalization of Formal Knowledge.* Chicago, IL: University of Chicago Press, 1986.

Hacker, Diana. *A Writer's Reference.* 4th ed. New York: St. Martin's Press, 1999.

Horowitz, Irving L., ed. *The Use and Abuse of Social Science: Behavioral Science and Policy Making.* 2nd ed. New Brunswick, NJ: Transaction Books, 1985.

Jasanoff, Sheila. *The Fifth Branch: Science Advisors as Policymakers.* Cambridge, MA: Harvard University Press, 1990.

Lindblom, Charles E., and David K. Cohen. *Usable Knowledge: Social Science and Social Problem Solving.* New Haven, CT: Yale University Press, 1979.

Machlup, Fritz. *Knowledge: Its Creation, Distribution, and Economic Significance*. Vol. 1, *Knowledge and Knowledge Production.* Princeton, NJ: Princeton University Press, 1980.

McCloskey, Deidre. *Economical Writing*. 2nd ed. Long Grove, IL: Waveland Press, 2000.

Rothman, Jack. *Social R&D: Research and Development in the Human Services*. Englewood Cliffs, NJ: Prentice Hall, 1980.

Tufte, Edward R. *The Visual Display of Quantitative Information.* Cheshire, CT: Graphics Press, 1983.

Tufte, Edward R. *Envisioning Information.* Cheshire, CT: Graphics Press, 1990.

Tufte, Edward R. *Visual Explanations.* Cheshire, CT: Graphics Press, 1997.

Weiss, Carol H. with Michael Bucuvalas. *Social Science Research and Decision Making.* New York: Columbia University Press, 1980.

CASE 9.1 TRANSLATING POLICY ARGUMENTS INTO ISSUE PAPERS, POSITION PAPERS, AND LETTERS TO THE EDITOR—THE CASE OF THE SECOND PERSIAN GULF WAR

An important skill is the translation of policy arguments into structured written narratives. These written narratives go by different names: white papers, issue papers, memos, press releases, and letters. All are used to communicate the results of analysis in hopes of maximizing the likelihood that it will actually be used to make and influence decisions.

In Chapter 8, the case of the Second Persian Gulf War was presented. The argument map showing the pros and cons of United States intervention was shown in Figure C8.2.1.

This argument map can be translated into four kinds of policy documents:

- *The Position Paper*. The position paper states only the reasons for making a policy claim. No opposing points of view are included; objections and rebuttals are not included. Position papers state one side of an argument as persuasively as possible.
- *The Policy Brief*. The issue brief is similar to a legal brief. It states the reasons for a claim, followed by objections, which are rebutted. Issue papers and memos state an opposing argument and then show why it does not hold up.
- *The Policy Issue Paper or Policy Memo*. The policy issue paper or policy memo (a shortened version of the policy issue paper) presents both sides of an issue equally, without taking sides. Reasons and objections are provided to support disputed claims or issues.
- *The Letter to the Editor*. The letter to the editor begins by considering an objection to the claim, rebutting it, and then providing multiple reasons to support the claim. The letter to the editor is an example of *a fortiori* analysis described in Chapter 5.

In each of the following written documents, the original argument map has been translated into a structured written narrative. The statements in **italics** letters are taken directly from the argument map. The documents would require editing before release.

Narrative for Position Paper on the U.S. Invasion of Iraq

There are three reasons to believe that the Senate should authorize the use of force under the Dole-Warner Resolution.

First, the Dole-Warner Resolution is justified by U.N. Security Council Resolution 678 of November 29, 1990. This is evident given that

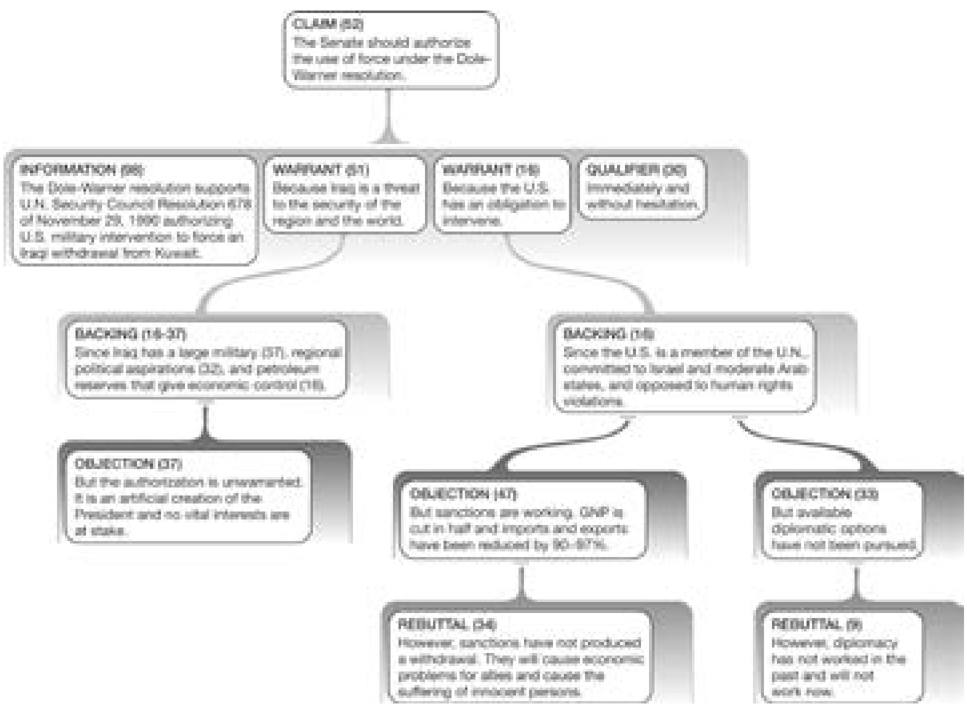


FIGURE C9.1.1
Sample Argument Map

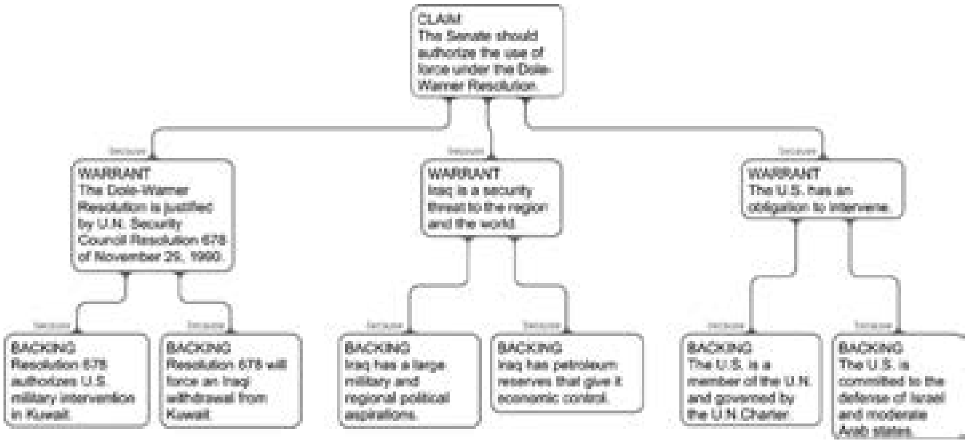


FIGURE C9.1.2
Argument Map for a Position Paper

Resolution 678 authorizes U.S. military intervention in Kuwait. Further evidence is that Resolution 678 will force an Iraqi withdrawal from Kuwait.

Second, Iraq is a security threat to the region and the world. This is supported by the reason that Iraq has large military and regional political aspirations. Furthermore, Iraq has petroleum reserves that give it economic control.

Finally, the U.S. has an obligation to intervene. Evidence for this is that the U.S. is a member of the

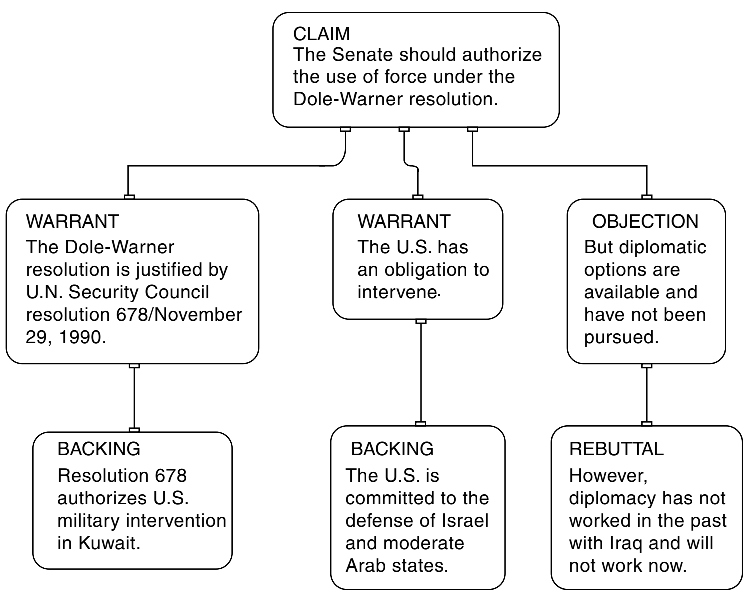


FIGURE C9.1.3
Argument Map for Policy Brief on the Invasion of Iraq

U.N. and governed by the U.N. Charter. In addition, the U.S. is committed to the defense of Israel and moderate Arab states, which also shows that the U.S. has an obligation to intervene.

Given these considerations, it is clear that the Senate should authorize the use of force under the Dole-Warner Resolution.

Narrative for Policy Brief

The Senate should authorize the use of force under the Dole-Warner Resolution. There are two main reasons supporting this contention, and the main objection to it can be rebutted.

The first reason is that the Dole-Warner Resolution is justified by U.N. Security Council Resolution 678 of November 29, 1990. This reason is based on the claim that Resolution 678 authorizes U.S. military intervention in Kuwait.

The second reason to think that the Senate should authorize the use of force under the Dole-Warner Resolution is that the U.S. has an obligation to intervene. This is because the U.S. is committed to the defense of Israel and moderate Arab states.

On the other hand, a consideration against the idea that the Senate should authorize the use of force under the Dole-Warner Resolution is that diplomatic options are available and have not been pursued. This objection is not convincing, however, because diplomacy has not worked in the past with Iraq and will not work now.

Based on this reasoning it is clear that the Senate should authorize the use of force under the Dole-Warner Resolution.

Narrative for Policy Memo on the U.S. Invasion of Iraq

The question has been raised whether *the Senate should authorize the use of force under the Dole-Warner Resolution.* This analysis considers a number of perspectives to determine whether to support or reject this claim.

One reason to accept that the Senate should authorize the use of force under the Dole-Warner Resolution is the suggestion that the Dole-Warner Resolution is justified by U.N. Security Council Resolution 678 of November 29, 1990. This reason is supported by the fact that Resolution 678 authorizes U.S. military intervention in Kuwait. An objection to the claim that the Dole-Warner Resolution is justified by U.N. Security Council Resolution 678 of November 29, 1990 is that the resolution is an artificial creation of the President and no vital interests are at stake.

Another reason to believe that the Senate should authorize the use of force under the Dole-Warner Resolution is that Iraq is a security threat to the region and the world. This is based on the idea that Iraq has large military and regional political aspirations.

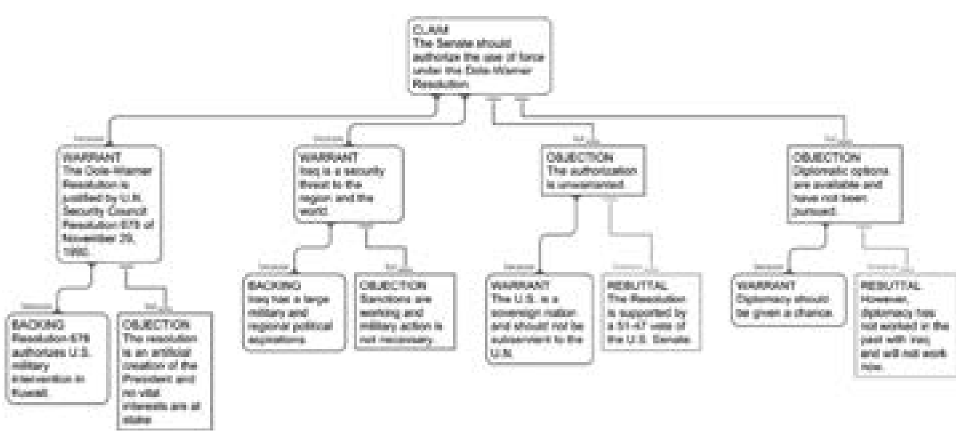


FIGURE C9.1.4
Argument Map for a Policy Issue Paper or Policy Memo on the Invasion of Iraq

Nonetheless, the opposing view that *Sanctions are working and military action is not necessary* provides an interesting perspective. On balance, however, it seems more reasonable to reject this argument.

On the other hand, a consideration against the claim that the Senate should authorize the use of force under the Dole-Warner Resolution is that the authorization is unwarranted. This objection is supported by the claim that the U.S. is a sovereign nation and should not be subservient to the U.N. Nevertheless, the rebuttal that the Resolution is supported by a 51–47 vote of the U.S. Senate must be considered to determine whether the objection is indeed acceptable.

Finally, there is another significant objection to the claim that the Senate should authorize the use of force under the Dole-Warner Resolution. Diplomatic options are available and have not been pursued. To support this view is the reason that diplomacy should be given a chance. In rebuttal, it is suggested that diplomacy has not worked in the past with Iraq and will not work now.

In summary, it would appear more reasonable to accept the claim that the Senate should authorize the use of force under the Dole-Warner Resolution. The strongest line of reasoning for this view is that Iraq is a security threat to the region and the world. Iraq has large military and political aspirations.

Narrative for a Letter to the Editor on the U.S. Invasion of Iraq

Dear Sir/Madam,

I am writing to contend that the Senate should authorize the use of force under the Dole-Warner Resolution. This is contrary to the view voiced by the Senate minority, which has argued that sanctions are working and military action is not necessary. I reject this claim because diplomacy has not worked in the past with Iraq, and it will not work now.

Furthermore, there are three strong reasons to support my view that the Senate should authorize the use of force under the Dole-Warner Resolution. The first reason is that the Dole-Warner Resolution is justified by U.N. Security Council Resolution 678 of

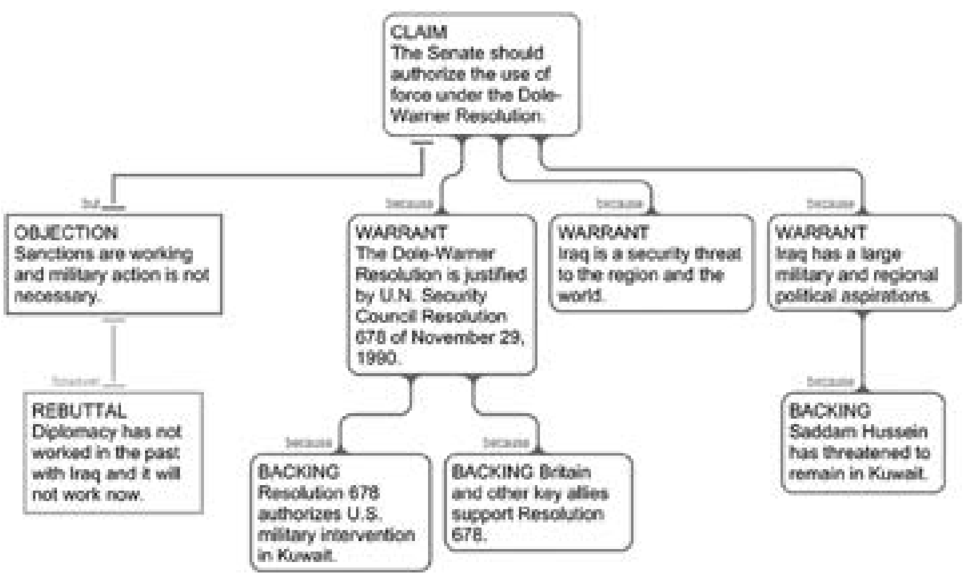


FIGURE C9.1.5
Argument Map for a Letter to the Editor

November 29, 1990, since Resolution 678 authorizes U.S. military intervention in Kuwait. In addition, this reason is supported by the fact that Britain and other key allies support Resolution 678.

The second reason is that Iraq is a security threat to the region and the world.

Finally, it is clear that the Senate should authorize the use of force under the Dole-Warner Resolution, given that Iraq has a large military and regional political aspirations. This reason is supported by the contention that Saddam Hussein has threatened to remain in Kuwait.

I restate my contention that the Senate should authorize the use of force under the Dole-Warner Resolution. I urge you to also adopt this view.

Yours sincerely,

[Your name] ■

CASE 9.2 COMMUNICATING COMPLEX ANALYSES TO MULTIPLE AUDIENCES: REGULATING LEADED GASOLINE

When Statistics Count: Revising the EPA Lead Standard[22]

David L. Weimer and Aidan L. Vining

Policy analysts must deal with many kinds of empirical evidence. Often the constraints of time and resources, as well as the nature of the policy problem under consideration, force analysts to rely on qualitative and fragmentary data. Sometimes, however, analysts can find data that enable them to estimate the magnitudes of the effects of policy interventions on the social, economic, or political conditions. The estimates may permit the analysts to calculate the likely net social benefits of proposed policies or even apply formal optimization techniques to find better alternatives.

The effective use of quantitative data requires an understanding of the basic issues of research design and a facility with the techniques of statistical inference. Even when analysts do not have the resources available to do primary data analysis, they often must confront quantitative evidence produced by other participants in the policy process or extracted from the academic literature. If they lack the requisite skills for critical evaluation, they risk losing their influence in the face of quantitative evidence that may appear more objective and scientific to decision-makers. Well-trained analysts, therefore, need some facility with the basic concepts of statistical inference for reasons of self-defense, if for no others.

The basics of research design and statistical inference cannot be adequately covered in an introductory course in policy analysis. Similarly, we cannot provide adequate coverage in this book. Yet we can provide an example of a policy change where quantitative analysis was instrumental: the decision by the U.S. Environmental Protection Agency in 1985 to reduce dramatically the amount of lead permitted in gasoline. The story of the new lead standard has a number of elements frequently encountered when doing quantitative analysis in organizational settings: replacement of an initial "quick and dirty" analysis with more sophisticated versions as more time and data become available; repeated reanalysis to rule out alternative explanations (hypotheses) offered by opponents of the proposed policy; and serendipitous events that influence the analytical strategy. Although our primary purpose is to tell what we believe to be an

[22]From David Weimer and Aidan Vining, *Policy Analysis: Concepts and Practice,* 2nd ed. (Englewood Cliffs, NJ: Prentice Hall Publishers, 1992), ch. 13, pp. 382–406. The remaining footnotes in this chapter originate from Weimer and Vining.

intrinsically interesting and instructive story about the practice of policy analysis, we hope to offer a few basic lessons on quantitative analysis along the way.

Background: The EPA Lead Standards

The Clean Air Amendments of 1970 give the administrator of the Environmental Protection Agency authority to regulate any motor fuel component or additive that produces emission products dangerous to public health or welfare.[23] Before exercising this authority, however, the administrator must consider "all relevant medical and scientific evidence available to him" as well as alternative methods that set standards for emissions rather than for fuel components.[24]

In 1971 the EPA administrator announced that he was considering possible controls on lead additives in gasoline.[25] One reason given was the possible adverse health effects of emissions from engines that burned leaded gasoline, the standard fuel at the time. Available evidence suggested that lead is toxic in the human body, that lead can be absorbed into the body from ambient air, and that gasoline engines account for a large fraction of airborne lead. The other reason for controlling the lead content of gasoline was the incompatibility of leaded fuel with the catalytic converter, a device seen as having potential for reducing hydrocarbon emissions from automobiles. The first reason suggests considering whether reductions in the lead content of all gasoline might be desirable; the second argues for the total removal of lead from gasoline for use by new automobiles equipped with catalytic converters.

Without going any further one might ask: Why should the government be concerned with the level of lead in gasoline? And why had the federal government already decided to require the installation of catalytic converters? The following analysis does not explicitly deal with the desirability of catalytic converters. As we have argued, however, it is always important to make sure that there is a convincing rationale for public action. The interventions at hand, the catalytic converter and the related lead restrictions, deal with market failures—viewed as problems of either negative externalities or public goods (ambient public goods). In addition, another market failure may come into play with respect to lead—imperfect consumer information about the impacts of lead on health and the maintenance cost of vehicles. Because the health and maintenance impacts may not manifest themselves for many years (leaded gasoline is a post-experience good), markets may be inefficient because of information asymmetry. These apparent market failures make a case for considering government intervention.

But intervention itself may be costly. It is quite conceivable that the costs of intervention could exceed the benefits. Several questions must be answered before the comparison of costs and benefits can be made: What are the impacts in the Economic Analysis Division who had worked on the 1982 regulations. Schwartz used his accumulated knowledge from previous analyses to complete in two or three days a quick and dirty benefit–cost analysis of a total ban on lead additives.

Because a total ban was one of the options considered in 1982, it was fairly easy for Schwartz to come up with a reasonable estimate of cost. As is often the case in evaluating health and safety regulations, the more difficult problem was estimating the benefits of a total ban. Schwartz looked at two benefit categories: increased IQ scores of children due to lower levels of blood lead and the avoided damage to catalytic converters.

Schwartz used estimates from a variety of sources to piece together a relationship between lead emissions and the present value of lifetime earnings of children. The first step involved using estimates from the 1982 analyses of the relationship between lead emissions and blood lead levels in children. He next turned to epidemiological studies that reported a relationship between blood lead levels and IQs. Finally, he found econometric studies that estimated the contribution of IQ points to the present value of future earnings.

As a first cut at quantifying the benefits resulting from the more effective control of other emissions,

[23]The Clean Air Act Amendments of 1970, Public Law 91–604, December 31, 1970.
[24]Section 211(c)(2)(A), 42 U.S.C. @ 1857f-6c(c)(2)(A).
[25]36 *Fed. Reg.* 1468 (January 31, 1971).

Schwartz estimated the cost of catalytic converters being contaminated under the current standards that would not be contaminated under a total ban. He used the number of converters saved multiplied by the price per converter as the benefit measure. Assuming that converters themselves have benefit–cost ratios greater than one, this benefit measure would be conservative.

These "back-of-the-envelope" calculations suggested that the benefits of a total ban on lead additives would be more than twice the costs. Schwartz discussed the results with his branch chief, G. Martin Wagner, who sent word to the office of the administrator that further analysis of a lead ban seemed worthwhile. A few weeks later Schwartz and fellow analyst Jane Leggett began a two-month effort to move from the back of an envelope to a preliminary report.

Pulling the Pieces Together

The most urgent task facing Schwartz and Leggett was to develop better measures of the benefits of a lead ban. The key to improving the measure of benefits from avoided converter poisonings was a more sophisticated accounting of the future age composition of the U.S. vehicle fleet. The key to improving the measure of benefits from reduced lead emissions was a better quantitative estimate of the relationship between gasoline lead and blood lead. Modeling and statistical analysis would be an important part of their work.

The age composition of the vehicle fleet continually changes as old vehicles, some of which have contaminated or partially contaminated converters, are retired and new ones, with fresh converters, are added. A ban on lead would have different effects on different vintage vehicles. For instance, it would be irrelevant for vehicles that already have contaminated converters, but very important for vehicles that would otherwise be contaminated and remain in service for a long time.

The analysts developed an inventory model that tracked cohorts of vehicles over time. Each year a new cohort would enter the fleet. Each successive year a fraction of the vehicles would be retired because of accidents and mechanical failures. In addition, a fraction would suffer converter poisoning from misfueling. By following the various cohorts over time, it was possible to predict the total number of converters that would be saved in each future year from specified reductions in the lead content of leaded gasoline today and in the future. Avoided loss of the depreciated value of the catalytic converters (later avoided costs of the health and property damage from pollutants other than lead) could then be calculated for each future year. With appropriate discounting, the yearly benefits could then be summed to yield the present value of the lead reduction schedule being considered.

Two important considerations came to light during work on the vehicle fleet model. One was a body of literature suggesting that lead increases the routine maintenance costs of vehicles. Subsequently, a new benefit category, avoided maintenance costs, was estimated using the vehicle fleet model. The other consideration was the possibility that some engines might suffer premature valve-seat wear if fueled with totally lead-free gasoline. Although the problem was fairly limited (primarily automobile engines manufactured prior to 1971 and some newer trucks, motorcycles, and off-road vehicles), it suggested the need to consider alternatives to a total ban.

Efforts to quantify better the link between gasoline lead and blood lead focused on analysis of data from the second National Health and Nutrition Examination Survey (NHANES II). The NHANES II survey was designed by the National Center for Health Statistics to provide a representative national sample of the population between ages 6 months and 74 years. It included 27,801 persons sampled at sixty-four representative sites from 1976 to 1980. Of the 16,563 persons who were asked to provide blood samples, 61 percent complied. The lead concentrations in the blood samples were measured, providing data that could be used to track average blood levels over the 4-year survey period. The amount of lead in gasoline sold over the period could then be correlated with the blood lead concentrations.

A positive relationship between blood lead concentrations from the NHANES II data and gasoline lead had already been found by researchers. The positive correlation is apparent in Figure C9.2.1, which was prepared by James Pirkle of the Centers for Disease Control to show the close tracking of blood lead concentrations and total gasoline lead.

Much more than the apparent correlation, however, was needed for the benefit–cost analysis. Schwartz used multiple regression techniques, which we will discuss in detail later in our story, to estimate the increase in average micrograms of lead per deciliter of blood (μg/dl) due to each additional 100 metric tons per day of lead in consumed gasoline. He also employed models that allowed him to estimate the probabilities that children with specified characteristics will have toxic levels of blood lead (then defined by the Centers for Disease Control as greater than 30 μg/dl) and gasoline lead. These probabilities could then be used to predict the number of children in the population who would avoid lead toxicity if a ban on gasoline lead were imposed.

In early November of 1983, Schwartz and Leggett pulled together the components of their analyses and found a ratio of benefits to costs of a total ban to be greater than Schwartz's original back-of-the-envelope calculation. Together with their branch chief, they presented their results to Deputy Administrator Alm, who found them encouraging. He gave the green light for the preparation of a refined version that could be presented to the administrator as the basis for a new regulation. Alm also wanted the various components of the refined study to be subjected to peer review by experts outside the EPA. At the same time, he urged speed in order to reduce the chances that word would get out to refiners and manufacturers of lead additives, the primary opponents of a ban, before the EPA had an opportunity to review all the evidence.

The branch chief expanded the analytical team in order to hasten the preparation of the refined report that would be sent to the administrator. Joining Schwartz and Leggett were Ronnie Levin; Hugh Pitcher, an econometrician; and Bart Ostro, an expert on the benefits of ozone reduction. In little more than one month the team was ready to send a draft report out for peer review.

The effort involved several changes in analytical approach. Because of the valve-head problem, the analysis focused on a major reduction in grams of lead per leaded gallon (from 1.1 gplg to 0.1 gplg) as well as a total ban. The available evidence suggested that the 0.1 gplg level would be adequate to avoid excessive valve-head wear in the small number of engines designed to be fueled only with leaded gasoline. At the same time, considerable effort was put into quantifying the maintenance costs that would be avoided by owners of other vehicles if the lead concentration were reduced. It soon appeared that the maintenance benefits consumers would enjoy would more than offset the higher prices they would have to pay for gasoline. Finally, the team decided that the benefits based on the blood lead (through IQ) to future earnings relationship would be too controversial. Instead, they turned their attention to the costs of compensatory education for children who suffered IQ losses from high blood lead levels.

In late December, sections of the report were sent to experts outside the EPA for comments. The list included automotive engineers, economists, biostatisticians, toxicologists, clinical researchers, transportation experts, and a psychologist. During January 1984, the team refined their analysis and incorporated, or at least responded to, the comments made by the external reviewers.

Finally, in early February, the team was ready to present its results to Administrator Ruckelshaus. He agreed that their analysis supported a new standard of 0.1 gplg. He told the team to finalize their report without a proposed rule and release it for public comment. Ruckelshaus also directed the Office of the Assistant Administrator for Air and Radiation to draft a proposed rule.

The team's *Draft Final Report* was printed and eventually released to the public on March 26, 1984.[26] The team continued to refine the analysis in the following months. It also had to devote considerable time to external relations. Executive Order 12291 requires regulatory agencies to submit proposed regulations that would have annual costs of more than $100 million to the Office of Management and Budget for review. The team met with OMB analysts several times before securing

[26]Joel Schwartz, Jan Leggett, Bart Ostro, Hugh Pitcher, and Ronnie Levin, *Costs and Benefits of Reducing Lead in Gasoline: Draft Final Report* (Washington, DC: Office of Policy Analysis, EPA, March 26, 1984).

their acceptance of the benefit–cost analysis of the tighter standard.

The political environment was taking shape pretty much as expected. Opposition would come from refiners and manufacturers of lead additives. The refiners, however, generally seemed resigned to the eventual elimination of lead from gasoline. Their primary concern was the speed of implementation. Some refiners seemed particularly concerned about the first few years of a tighter standard, when they would have difficulty making the required reductions with their existing configurations of capital equipment. In response, the team began exploring the costs and benefits of less stringent compliance schedules.

The manufacturers of lead additives were ready to fight tighter standards. In May Schwartz attended a conference at the Centers for Disease Control in Atlanta, Georgia, on the proposed revision of the blood lead toxicity standard for children from 30 μg/dl to 25 μg/dl. Representatives of the lead manufacturers were also there. They openly talked about their strategy for challenging tighter standards. They planned to argue that refiners would blend more benzene, a suspected carcinogen, into gasoline to boost octane if lead were restricted. Schwartz investigated this possibility following the conference. He found that, even if more benzene were added to gasoline, the total emissions of benzene would decline because of the reduction in the contamination of catalytic converters, which oxidize the benzene if not contaminated. The day that the proposed rule was published Schwartz put a memorandum on the docket covering the benzene issue, thus preempting the manufacturers' main attack.

The EPA published the proposed rule on August 2, 1984.[27] It would require that the permitted level of gasoline lead be reduced to 0.1 gplg on January 1, 1986. The proposal stated the EPA assumption that the new standard could be met with existing refining equipment, but indicated that alternative phase-in schedules involving more gradual reductions also were being considered in case this assumption proved false. Finally, the proposal raised the possibility of a complete ban on gasoline lead by 1995.

A Closer Look at the Link between Gasoline Lead and Blood Lead

Calculation of the direct health benefits of tightening the lead standard requires quantitative estimates of the contribution of gasoline lead to blood lead. The NHANES II data, combined with information on gasoline lead levels, enabled the study team to make the necessary estimates. Their efforts provide an excellent illustration of how statistical inference can be used effectively in policy analysis.

The Need for Multivariate Analysis

A casual inspection of Figure C9.2.1 suggests a strong positive relationship between gasoline lead and blood lead. Why is it necessary to go any further? One reason is the difficulty of answering, directly from Figure C9.2.1, the central empirical question: How much does the average blood lead level in the United States decline for each 1,000-ton reduction in the total gasoline lead used over the previous month? Figure C9.2.1 indicates a positive correlation between gasoline lead and blood lead—changes in blood lead track changes in gasoline lead quite closely. But for the same correlation in the data, we could have very different answers to our central question.

Figure 9.3 illustrates the difference between correlation and magnitude of effect with stylized data. If our data were represented by the diamonds, we might "fit" line one as our best guess at the relationship between blood lead and gasoline lead. The effect of reducing gasoline lead from 500 tons per day to 400 tons per day would be to reduce average blood lead from 10.1 μg/dl to 10.0 μg/dl, or 0.1 μg/dl per 100 tons per day. An alternative sample of data is represented by the dots, which are plotted to have approximately the same correlation between gasoline lead and blood lead as the data represented by triangles. The slope of line two, the best fit to the dots, is 1.0 μ g/dl per 100 tons per day—ten times greater than the slope of line one. That is, although the two data sets have the same correlation, the second implies a much greater contribution of gasoline lead to blood lead than the first.

[27] 49 *Fed. Reg.* 31031 (August 2, 1984).

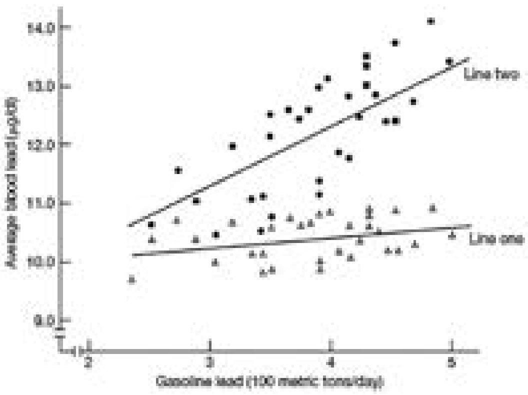


FIGURE 9.3
Data samples with correlations but different regression lines

Even after we display the data in Figure C9.2.1 as a plot between blood lead and gasoline lead, our analysis is far from complete because the apparent relationship (the slopes of lines one and two in our example) may be spurious. Blood lead and gasoline lead may not be directly related to each other but to some third variable that causes them to change together. The classic illustration of this problem is the correlation sometimes found between the density of stork nests and the human birth rate. If one were to plot birth rates against the density of stork nests for districts in some region, one might very well find a positive relationship and perhaps conclude that there might be something to the myth about storks bringing babies.

Of course, there is a more plausible explanation. A variable measuring the degree to which a district is rural "intervenes" between birth rate and nest density—rural areas have farmers who want a lot of children to help with chores as well as lot of open land that provides nesting grounds for storks; more urbanized areas tend to have people who want smaller families as well as less hospitable nesting grounds. Looking across a sample that included both rural and urban districts could yield the positive correlation between birth rate and nest density. If we were to "control" statistically for the intervening variable by looking at either rural or urban districts separately rather than pooled together, we would expect the correlation between birth rate and nest density to become negligible.

Returning to our analysis of lead, we must be concerned that one or more intervening variables might explain the positive correlation between gasoline lead and blood lead. For example, there is some evidence that cigarette smokers have higher blood lead levels than nonsmokers. It may be that over the period that the NHANES II data were collected the proportion of smokers in the sample declined, so that much of the downward trend in average blood lead levels should be attributed to reductions in smoking rather than reductions in gasoline lead.

To determine whether smoking is an intervening, or confounding, variable, we might construct separate diagrams like Figure 9.3 for smokers and nonsmokers. In this way, we could control for the possibility that changes in the proportion of smokers in the sample over time were responsible for changes in blood lead levels. If we fit lines with similar positive slopes to each of the samples, we would conclude that smoking behavior was not an intervening variable for the relationship between gasoline lead and blood lead.[28]

With unlimited quantities of data, we could always control for possible intervening variables in this way. Unfortunately, with a fixed sample size, we usually stretch our data too thin if we try to form subsets of data that hold all variables constant (e.g., everyone is a nonsmoker) except the dependent variable (blood lead levels) we are trying

[28]If the two lines coincided, we would conclude that smoking did not contribute to blood lead. If the lines were parallel but not identical, the difference between their intercepts with the vertical axis would represent the average effect of smoking on blood lead. If the lines were not parallel, we might suspect that smoking interacted with exposure to gasoline lead so that their effects are not additive. That is, smokers were either more or less susceptible to exposure to gasoline lead than nonsmokers.

[29]If the relationship truly varies across our subsets, then it would generally not make sense to make an overall estimate. As we will explain later, however, we never really observe the true relationship—it always contains some unknown error. Therefore, we may not be sure if it is reasonable to combine our subsample data. Because the variance of the error will be larger, the smaller the size of our subsamples, the more we divide our data, the more difficult it is to determine whether observed differences reflect true differences.

to explain and the independent variable (gasoline lead) we are considering as a possible explanation. For example, occupational exposure, alcohol consumption, region, and age are only a few of the other variables that might be causing the relationship in our data between gasoline lead and blood lead. If we selected from our sample only adult males living in small cities in the South who are moderate drinkers and nonsmokers and have no occupational exposure to lead, we might have too few data to reliably fit a line as in Figure 9.3. Even if we had adequate data, we would end up with estimates of the relationship between gasoline lead and blood lead from all of our subsets. (The number of subsets would equal the product of the number of categories making up our control variables.) We might then have difficulty combining these subset estimates into an overall estimate.[29]

The Basic Linear Regression Model

Linear regression provides a manageable way to control statistically for the effects of several independent variables.[30] Its use requires us to assume that the effects of the various independent variables are additive. That is, the marginal effect on the dependent variable of a unit change in any one of the independent variables remains the same no matter what the values of the other independent variables.[31] We can express a linear regression model in mathematical form:

$$y = b_0 + b_1x_1 + b_2x_2 + b_2x_2 + \ldots + b_kx_k + e$$

where y is the dependent variable, $x_1, x_2, \ldots, x_k$ are the k independent variables, $b_0, b_1, \ldots, b_k$ are parameters (coefficients) to be estimated, and e is an error term that incorporates the cumulative effect on y of all the factors not explicitly included in the model. If we were to increase x_1 by one unit while holding the values of the other independent variables constant, y would change by an amount b_1. Similarly, each of the coefficients measures the marginal effect of a unit change in its variable on the dependent variable.

Imagine that we set the values of all the independent variables except x_1 equal to zero. We could then plot y against x_1 in a graph like Figure 9.3. The equation $y = b_0 + b_1x_1$ would represent the line fit to our sample of observations. The slope of the line is b_1, the magnitude of the change in y that will result from a unit change in x_1, other things equal. The actual observations do not lie exactly on the line, however. Their vertical distances from the line will be equal to the random error, represented in our model by e, incorporated in

[30]For clear introductions, see Eric A. Hanushek and John E. Jackson, *Statistical Methods for the Social Sciences* (New York: Academic Press, 1977); and Christopher H. Achen, *Interpreting and Using Regression* (Beverly Hills, Calif.: Sage Publications, 1982).

[31]This assumption is not as restrictive as it might at first seem. We can create new variables that are functions of the original independent variables to capture nonlinearities. For example, if we thought smokers were likely to absorb environmental lead faster than nonsmokers, we might include in our linear regression model a new variable that is the product of the number of cigarettes smoked per day and the level of gasoline lead. This new variable, capturing the interaction of smoking and gasoline lead, would have an effect that would be additive with the other independent variables in the model, including the smoking and gasoline lead variables. The marginal effect of gasoline lead on blood lead would consist of the contribution from the gasoline lead variable and the contribution from the variable representing its interaction with smoking, which will change depending on the particular level of smoking.

[32]The *prediction error* is the observed value of the dependent variable minus the value we would predict for the dependent variables based on our parameter estimates and the values of our independent variables. For the *i*th observation the prediction error is given by:

$$y_i - (\hat{b}_0 + \hat{b}_1x_{1i} + \ldots + \hat{b}_kxk_i)$$

The prediction error is also referred to as the *residual of the observation.* OLS selects the parameter estimates to minimize the sum of the squares of the residuals.

[33]When one independent variable can be written as a linear combination of the others, we have a case of *perfect multicollinearity.* A related and more common problem, which we can rarely do anything about, occurs when the independent variables in our

each of the observations of y. If the values of e are small, our line will fit the data well in the sense that the actual observations will lie close to it.

How should we go about fitting the line? The most commonly used procedure is the method of *ordinary least squares* (OLS). When we have only one independent variable, so we can plot our data on a two-dimensional graph like Figure 9.3, the OLS procedure picks the line for which the sum of squared vertical deviations from the observed data is smallest. When we have more than one independent variable, OLS determines the values of the coefficients ($b_0, b_1, \ldots, b_k$) that minimize the sum of squared prediction errors.[32] As long as the number of observations in our sample exceeds the number of coefficients we are trying to estimate, and none of our independent variables can be expressed as a linear combination of the other independent variables, the commonly available regression software packages will enable us to use computers to find the OLS fitted coefficients.[33]

The estimates of coefficients that we obtain from OLS will generally have a number of very desirable properties. If the independent variables are all uncorrelated with the error term (e), then our coefficient estimators will be *unbiased*.[34] (*Note:* an *estimator* is the formula we use to calculate a particular estimate from our data.) To understand what it means for an estimator to be unbiased, we must keep in mind that our particular estimate depends upon the errors actually realized in our sample of data. If we were to select a new sample, we would realize different errors and hence different coefficient estimates. When an estimator is unbiased, we expect that the average of our estimates across different samples will be very close to the true coefficient value. For example, if gasoline lead had no true effect on blood lead, we would almost certainly estimate its coefficient to be positive or negative, rather than exactly zero. Repeating OLS on a large number of samples and averaging our estimates of the coefficient of gasoline lead, however, would generally yield a result very close to zero.[35] Indeed, by adding more and more samples, we could get the average as close to zero as we wanted.

Unfortunately, we usually only have a single sample for estimating coefficients. How do we decide if an estimate deviates enough from zero for us to

---

sample are highly correlated. This condition, called *multicollinearity,* is not a problem with the specification of our model but with the data we have available to estimate it. If two variables are highly correlated, positively or negatively, OLS has difficulty identifying their independent effects on the dependent variable. As a result, the estimates of the parameters associated with these variables will not be very reliable. That is, they will have large variances, increasing the chances that we will fail to recognize, statistically speaking, their effects on the dependent variable. One way to deal with multicollinearity is to add new observations to our sample that lower the correlation. For example, if we had a high positive correlation between smoking and drinking in our sample, we should try to add observations on individuals who smoke but do not drink and who drink but do not smoke. Unfortunately, we often have no choice but to work with the data that are already available.

[34]Strictly speaking, we must also assume that our independent variables are fixed in the sense that we could construct a new sample with exactly the same observations on the independent variables. Of course, even if the independent variables are fixed, we would observe different values of the dependent variable because of the random error term. In addition, for complete generality, we must assume that the expected value of the error term is constant for all observations.

[35]Our average will not be close to zero if gasoline lead is correlated with a variable excluded from our model that does have an effect on blood lead. In this case, gasoline lead stands as a proxy for the excluded variable. Other things equal, the stronger the true effect of the excluded variable on blood lead and the higher the absolute value of the correlation between gasoline lead and the excluded variable, the greater will be the bias of the coefficient of gasoline lead. We might not worry that much about the bias if we knew that it would approach zero as we increase sample size. (If the variance of the estimator also approached zero as we increased sample size, we would say that the estimator is *consistent.*) Although OLS estimators are consistent for correctly specified models, correlation with an important excluded variable makes an estimator inconsistent.

[36]The *Central Limit Theorem* tells us that the distribution of the sum of independent random variables approaches the Normal distribution as the number in the sum becomes large. The theorem applies for almost any starting distributions—the existence of a finite variance is sufficient. If we think of the error term as the sum of all the many factors excluded from our model, and further, we believe that they are not systematically related to each other or the included variables, then the Central Limit Theorem suggests that the distribution of the error terms will be at least approximately Normal.

conclude that the true value of the parameter is not zero? Making the fairly reasonable assumption that the error term for each observation can be treated as a draw from a Normal distribution with constant variance, the OLS estimators will be distributed according to the Student-*t* distribution.[36] That is, we can interpret the particular numerical estimate of a coefficient as a draw from a random variable distributed as a Student-*t* distribution centered around the true value of the coefficient. (The OLS estimator is the random variable; the actual estimate based on our data is a realization of that random variable.)

Knowing the distribution of the OLS estimator enables us to interpret the *statistical significance* of our coefficient estimate. We determine statistical significance by asking the following question: How likely is it that we would observe a coefficient estimate as large as we did if the true value of the coefficient were zero? We answer this question by first assuming that the true value of the coefficient is zero (the null hypothesis) so that the distribution of our estimator is centered around zero. We then standardize our distribution to have a variance of one by dividing our coefficient estimate by an estimate of its standard error (a by-product of the OLS procedure). The resulting number, called the *t*-ratio, can then be compared to critical values in tabulations of the standardized Student-*t* distribution found in the appendix of almost any statistics text. For example, we might decide that we will reject the null hypothesis that the true value of the coefficient is zero if there is less than a 5 percent probability of observing a *t*-ratio (in absolute value sense) as large as we did if the null hypothesis is true. (The probability we choose puts an upward bound on the probability of falsely rejecting the null hypothesis.)[37] To carry out the test, we look in the standardized tabulations of the Student-*t* distribution for the critical value corresponding to 5 percent.[38] If the absolute value of our estimated *t*-ratio exceeds the critical value, then we reject the null hypothesis and say that our estimated coefficient is statistically significantly different from zero.

Fortunately, most regression software saves us the trouble of looking up critical values in tables by directly calculating the probability under the null hypothesis of observing a *t*-ratio as large as that estimated. To do a classical test of hypothesis on the coefficient, we simply see if the reported probability is less than the maximum probability of falsely rejecting the null hypothesis that we are willing to accept. If it is smaller, then we reject the null hypothesis.

Consider the regression results presented in Table C9.2.1. They are based on data from 6,534 whites in the NHANES II survey for whom blood lead measurements were made.[39] The dependent variable is the individual's blood lead level measured in μg/dl. The independent variables are listed under the heading

[37]Falsely rejecting the null hypothesis is referred to as Type I error. Failing to reject the null hypothesis when in fact the alternative hypothesis is true is referred to as Type II error. We usually set the probability of Type I error at some low level like 5 percent. Holding sample size constant, the lower we set the probability of Type I error, the greater the probability of Type II error.

[38]The Student-*t* distribution is tabulated by degrees of freedom. In the basic OLS framework, the degrees of freedom is the total number of observations minus the number of coefficients being estimated. As the degrees of freedom becomes larger, the Student-*t* distribution looks more like a standardized Normal distribution.

You should also note the difference between a one-tailed and two-tailed test. Because the standardized Student-*t* is a symmetric distribution centered on zero, a 5 percent test usually involves setting critical values so that 2.5 percent of area lies under each of the tails (positive and negative). A one-tailed test, appropriate when the null hypothesis is that the true coefficient value is zero or less than zero, puts the entire 5 percent in the positive tail.

[39]The analysts estimated similar models for blacks and for blacks and whites together. Their estimates of the coefficient for gasoline lead never deviated by more than 10 percent across the different samples. To conserve space, they reported in detail only their regression results for whites. They chose whites because it was the largest subgroup and because it preempted the assertion that the relationship between blood lead and gasoline lead was due to changes in the racial composition of the sample over time.

[40]R^2 is a measure of the *goodness of fit* of the model to the particular sample of data. It is the square of the correlation between the values of the dependent variable predicted by the model and the values actually observed. An R^2 of one would mean that the model perfectly predicted the independent variable for the sample; an R^2 of zero would mean that the model made no contribution to prediction.

"Effect." The independent variable of primary interest is the national consumption of gasoline lead (in hundreds of metric tons per day) in the month prior to the individual's blood lead measurement. The other independent variables were included in an effort to control statistically for other factors that might be expected to affect the individual's blood lead level. With the exception of the number of cigarettes smoked per day and the dietary factors (vitamin C, riboflavin, and so on), these other statistical controls are indicator, or "dummy," variables that take on the value one if some condition is met and zero otherwise. So, for example, if the individual is male, the variable "male" will equal one; if the individual is female, it will equal zero. The variables "vitamin C," "phosphorus," "riboflavin," and "vitamin A," which are included as proxy measures for dietary intake of lead, each measure dietary intake in milligrams. Otherwise, the other variables are intended to capture demographic, income, occupational exposure, drinking habit, and locational effects. The reported R^2 indicates that, taken together, the independent variables explain about 33 percent of the total variation in blood lead levels.[40]

The estimated coefficient for gasoline lead is 2.14 μg/dl of blood per 100 metric tons per day of national gasoline lead consumption. Dividing the coefficient estimate by its estimated standard error of 0.192 yields a *t*-ratio of about 11. The probability of observing a *t*-ratio this large or larger if the true value of the coefficient were actually zero is less than one chance in 10,000 (the 0.0000 entry in Table C9.2.1 under "*P*-value"). We would thus reject the null hypothesis in favor of the alternative hypothesis that gasoline lead does contribute to blood lead. In other words, we would say that gasoline lead has a *statistically significant effect* on blood lead.

After finding a statistically significant effect, the next question to ask is whether the size of the coefficient is *substantively significant*. That is, does the variable in question have an effect that is worth considering?[41] One approach to answering this question is to multiply the estimated coefficient by the plausible change in the independent variable that might occur. For example, by the end of the NHANES II survey, gasoline lead was being consumed at a rate of about 250 metric tons per day nationally. A strict policy might reduce the level to, say, 25 metric tons per day. Using the estimated coefficient of gasoline lead, we would expect a reduction of this magnitude to reduce blood lead levels on average by about 4.8 μg/dl (the reduction of 225 metric tons per day times the estimated coefficient of 2.14 μg/dl per 100 metric tons per day).

To get a better sense of whether a 4.8 μg/dl reduction is substantively important, we can look at the blood lead levels for representative groups at the 250 and 25 metric ton levels. For example, at the 250 metric ton level a nonsmoking (number of cigarettes equals zero), moderate drinking (drinker equals one; heavy drinker equals zero), nonoccupationally exposed (occupationally exposed equals zero), adult female (child, teenager, male, male teenager, and male adult equal zero), living in a large Northeastern city (Northeast equals one; small city, rural, South, and Midwest equal zero), with moderate income, a college degree, and a high-nutrition diet (low income equals zero; moderate income, educational level, vitamin C, phosphorus, riboflavin,

---

[41]The standard errors of the coefficient estimates decrease as sample size increases. Thus, very large samples may yield large *t*-ratios even when the estimated coefficient (and its true value) are small. We refer to the power of a statistical test as 1 minus the probability of failing to reject the null hypothesis in favor of the alternative hypothesis. Other things equal, larger sample sizes have greater power, increasing the chances that we will reject the null hypothesis in favor of alternatives very close to 0.

[42]For an introduction to logistic regression, see Hanushek and Jackson, *Statistical Methods,* pp. 179–216.

[43]The logistic regression model is written:
$P(Y) = e^Z/(1 = e^Z)$ where $P(Y)$ is the probability that condition Y holds, e is the natural base, and $Z = b_0 + b_1x_1 + \ldots + b_kx_k$ is the weighted sum of the independent variables $x_1, x_2, \ldots, x_k$. The coefficients, $b_0, b_1, \ldots, b_k$, are selected to maximize the probability of observing the data in the sample. Note that the marginal contribution of x_i to the value of the dependent variable is not simply b_i, as would be the case in a linear regression model. Rather, it is $b_i[1 - P(Y)]P(Y)$, which has its greatest absolute value when $P(Y) = 0.5$.

[44]The logistic regression model, unlike the linear regression model, always predicts probabilities that lie between zero and one (as they should).

and vitamin A equal one) would be expected to have a blood lead level of 10.6 μg/dl. We would expect the same person to have a blood lead level of only 5.8 μg/dl if the gasoline lead level were cut to 25 metric tons per day—a reduction of about 45 percent. Thus, the effect of gasoline lead on blood lead appears substantively as well as statistically significant.

The study team was especially interested in estimating the contribution of gasoline lead to blood lead in children. As a first cut, they developed a logistic regression model[42] for predicting the probability that a child between the ages of six months and eight years will have blood levels in excess of 30 μg/dl, the definition of lead toxicity used by the Centers for Disease Control at the time. *Logistic regression,* which assumes a nonlinear relationship between the dependent and the independent variables,[43] is usually more appropriate than linear regression when the dependent variable is dichotomous (*Y* equals one if the condition holds, *Y* equals zero if it does not).[44] The study team found a strong relationship in the NHANES II data between gasoline lead and the probability that a child has a toxic level of blood lead. In fact, they estimated that the elimination of gasoline lead would have reduced the number of cases of lead toxicity in the sample by 80 percent for children under eight years of age. The study team used logistic regression and other probability models to estimate how reductions in gasoline lead would change the number of children having various blood lead levels. These estimates were essential for their subsequent valuation of the effects of gasoline lead reductions on the health of children.

Reconsidering Causality

Finding that an independent variable in a regression model has a statistically significant coefficient does not by itself establish a causal relationship. That is, it does not guarantee that changes in the independent variable cause changes in the dependent variable. Even if the independent variable has no direct effect on the dependent variable, some other variable, not included in the model, may be correlated with both so as to produce an apparent relationship in the data sample (remember the apparent relationship between birth rates and the density of stork nests). Should the strong relationship between gasoline lead and blood lead be interpreted as causal?

The study team considered this question in detail. Although its demonstration was not legally necessary, they believed that adoption of the proposed rule would be more likely if a strong case for causality could be made. Their approach was to apply the criteria commonly used by epidemiologists to determine the likelihood of causality. Not all of the criteria are directly applicable outside of the health area. Nonetheless, the way the study team applied the criteria illustrates the sort of questioning that is valuable in empirical research. Therefore, we briefly review the six criteria that they considered.

Is the Model Biologically Plausible? The study team noted that lead can be absorbed through the lung and gut. They pointed out that gasoline lead, the major source of environmental lead, is emitted predominantly as respirable particulates in automobile exhaust. These particulates can be absorbed directly through the lungs. They also contaminate dust that can be inhaled through the lungs and absorbed through the gut. Therefore, they argued, it is biologically plausible that gasoline lead contributes to blood lead.

Biological plausibility is the epidemiological statement of a more general criterion: Is the model theoretically plausible? Prior to looking through data for empirical relationships, you should specify a

[45]Imagine that you regress a variable on twenty other variables. Assume that none of the twenty independent variables has an effect on the dependent variable. (The coefficients in the true model are all zero.) Nevertheless, if you use a statistical test that limits the probability of falsely rejecting the null hypothesis to 5 percent, then you would still have a 0.64 probability $[1 - (.95)^{20}]$ of rejecting at least one null hypothesis. In other words, if you look through enough data you are bound to find some statistically significant relationships, even when no true relationships exist. By forcing yourself to specify theoretical relationships before you look at the data, you reduce the chances that you will be fooled by the idiosyncrasy of your particular data sample. For a brief review of these issues, see David L. Weimer, "Collective Delusion in the Social Sciences: Publishing Incentives for Empirical Abuse," *Policy Studies Review,* 5, 4 (May 1986), pp. 705–708.

[46]S. Fachetti and F. Geiss, *Isotopic Lead Experiment Status Report,* Publication No. EUR8352ZEN (Luxembourg: Commission of the European Communities, 1982).

model (your beliefs about how variables are related). If you find that your data are consistent with your model, then you can be more confident that the relationships you estimate are not simply due to chance.[45]

Is there Experimental Evidence to Support the Findings? The study team found reports of several investigations specifically designed to measure the contribution of gasoline lead to blood lead. One was an experiment conducted in Turin, Italy, by researchers who monitored changes in the isotopic composition of blood lead as the isotopic composition of gasoline lead varied.[46] They found that at least 25 percent of the lead in the blood of Turin residents originated in gasoline. Thus, the experiment not only confirmed the biological plausibility of the contribution of gasoline lead to blood lead, but also suggested an effect on the order of magnitude of that estimated by the study team.

Being able to find such strong and directly relevant experimental support is quite rare in policy research, much of which deals with the behavioral responses of people. Controlled experiments in the social sciences are rare, not only because they are costly and difficult to implement, but also because they often involve tricky ethical issues concerning the assignment of people to "treatment" and "control" groups. Nevertheless, there have been a number of policy experiments in the United States over the last twenty years.[47] It is unlikely, however, that any of these experiments will be directly applicable to your policy problem. Therefore, you must typically broaden your search for confirmation beyond experiments to other empirical research.

Do Other Studies Using Different Data Replicate the Results? The study team reviewed several studies that also found relationships between gasoline lead and blood lead. These studies were based on data collected in conjunction with community-wide lead-screening programs funded by the Centers for Disease Control during the 1970s[48] and on data collected from the umbilical cord blood of over 11,000 babies born in Boston between April 1979 and April 1981.[49] These studies reported statistically significant relationships between gasoline lead and blood lead, and thus supported the study team's analysis based on the NHANES II data.

Does Cause Precede Effect? The study team used information about the half-life of lead in blood to make predictions about the strengths of relationships between lagged levels of gasoline lead and blood lead that would be expected if gasoline lead contributes to blood lead. Lead has a half-life of about thirty days in blood. Noting that the average NHANES II blood test was done at mid-month, they predicted that the previous month's gasoline lead (which on average represents emissions occurring between fifteen and forty-five days before the test) should have a stronger impact on blood lead than that of either the current month (average exposure of zero to fifteen days) or the month occurring two months prior (average exposure of forty-five to seventy-five days). They tested their predictions by regressing blood lead levels on current, one-month lagged, and two-month lagged gasoline lead levels. As predicted, one-month lagged gasoline lead was the most significant of the three. Also, consistent with the thirty-day half-life, two-month lagged gasoline lead had a coefficient approximately one-half that of one-month lagged gasoline lead. Thus, cause did appear to precede effect in the expected way.

Does a Stable Dose-Response Relationship Exist? The regression model used by the study team assumed a linear relationship between gasoline lead and blood lead. Did this relationship remain stable as the level of gasoline lead changed? To answer this question, the study team took advantage of the fact that, on average, gasoline lead levels were about 50 percent lower in the second half of the NHANES II survey than in the first half. If the relationship

---

[47]For a review of the major policy experiments conducted in the United States, see David H. Greenberg and Philip K. Robins, "The Changing Role of Social Experiments in Policy Analysis," *Journal of Policy Analysis and Management,* 5, 2 (Winter 1986), pp. 340–362.

[48]Irwin H. Billick et al., *Predictions of Pediatric Blood Lead Levels from Gasoline Consumption* (Washington, DC: U.S. Department of Housing and Urban Development, 1982).

[49]Michael Rabinowitz and Herbert L. Needleman, "Petrol Lead Sales and Umbilical Cord Blood Lead Levels in Boston, Massachusetts," *Lancet,* January 1/8, 1983, p. 63.

between gasoline lead and blood lead is stable and linear, then reestimating the regression model using only data from the second half of the survey should yield a coefficient for gasoline lead comparable to that for the entire sample. They found that the coefficients were indeed essentially the same. In addition, estimation of regression models that directly allowed for the possibility of nonlinear effects supported the initial findings of a linear relationship between gasoline lead and blood lead.

Is it Likely That Factors Not Included in the Analysis Could Account for the Observed Relationship? The study team considered several factors that might confound the apparent relationship between gasoline lead and blood lead: dietary lead intake, exposure to lead paint, seasonality, and sampling patterns.

The basic regression model included nutrient and demographic variables as proxy measures for the intake of lead in the diet. Yet these variables may not adequately control for a possible downward trend in dietary lead that could be causing the estimated relationship between gasoline lead and blood lead. Market basket studies conducted by the Food and Drug Administration over the survey period, however, showed no downward trend in dietary lead intake. Also, lead intake from drinking water is largely a function of acidity, which did not change systematically over the survey period. Evidence did suggest that changes in solder reduced the content of lead in canned foods over the period. But the study team was able to rule out the lead content in canned foods as a confounding factor when they added the lead content in solder as an independent variable, reestimated the basic regression model, and found that the coefficient of gasoline lead remained essentially unchanged.

The study team recognized changing exposure to lead paint as another potential confounding factor. They dismissed the possibility on three grounds.

First, paint lead is a major source in blood lead for children (who eat paint chips) but not for adults. If declining exposure to lead paint were responsible for the estimated relationship between gasoline lead and blood lead, we would expect the reduction in blood lead to be much greater for children than for adults. In fact, the average reduction over the survey period for adults was only slightly smaller than that for children (37 percent versus 42 percent).

Second, the ingestion of paint lead usually results in large increases in blood lead levels. If reduced exposure to paint lead were responsible for declining blood lead levels, then we would expect to observe the improvement primarily in terms of a reduction in the number of people with very high blood lead levels. In fact, blood lead levels declined over the survey period even for groups with low initial levels.[50]

Third, declining exposure to paint lead should be a more important factor in central cities than in suburbs because the latter tend to have newer housing stocks with lower frequencies of peeling lead paint. Yet the gasoline lead coefficient was essentially the same for separate estimations based on central city and suburban subsamples.

Blood lead levels in the United States are on average higher in the summer than in the winter. To rule out the possibility that seasonal variation confounds the relationship between gasoline lead and blood lead, the study team reestimated the basic model with indicator variables included to allow for the possibility of independent seasonal effects. The coefficients of the seasonal variables were not statistically significant when gasoline lead was kept in the model. Thus, it appeared that changes in gasoline lead could adequately explain seasonal as well as long-term changes in blood lead levels.

As already mentioned, the study team estimated the basic model on a variety of demographic subsamples and found no more than a 10 percent difference across any two estimates of the gasoline lead coefficient. They were also concerned, however, that changes in NHANES II sampling locations over the survey period might have confounded the estimation of the gasoline lead coefficient. Therefore, they reestimated the basic model with indicator variables for forty-nine locations and found that the gasoline lead coefficient changed by only about 5 percent. Further, they found that, even when including variables to allow for different gasoline

[50]The study team also used data from the lead-screening program in Chicago to estimate the probability of toxicity as a function of gasoline lead for children exposed and not exposed to lead paint. They found that gasoline lead had statistically significant positive coefficients for both groups.

lead coefficients across locations, the coefficient representing the nationwide effect of gasoline lead was statistically and substantively significant. Together, these tests led the study team to dismiss the possibility of serious sampling bias.

The Weight of the Evidence

The study team produced a very strong case in support of an important causal relationship between gasoline lead and blood lead. In many ways their efforts were exemplary. They drew relevant evidence from a wide variety of sources to supplement their primary data analysis. They gave serious attention to possible confounding factors, considering both internal tests (such as subsample analyses and model respecifications) and external evidence to see if they could be ruled out. As a consequence, opponents of the proposed policy were left with few openings for attacking its empirical underpinnings.

Finalizing the Rule

The primary task facing the analytical team after publication of the proposed rule was to respond to comments made by interested parties. Team members participated in public hearings held in August and spent much of the fall of 1984 responding to comments placed in the public docket, which closed on October 1. During this process they became more confident that the proposed rule would produce the large net benefits they predicted. At the same time they discovered another benefit category—reductions in adult blood pressure levels—that could potentially swamp their earlier estimates of benefits.

In 1983 Schwartz chanced upon a research article reporting a correlation between blood lead and hypertension.[51] He began work with researchers at the Centers for Disease Control and the University of Michigan to determine if a relationship existed between blood lead and blood pressure levels. By the summer of 1984 their analysis of the NHANES II data suggested a strong link.[52] Because high blood pressure contributes to hypertension, myocardial infarctions, and strokes, the potential benefits from blood lead reductions were enormous. Although the final rule was ultimately issued without reference to quantitative estimates of the benefits of lower adult blood lead levels, the team provided estimates in the supporting documents.

The one remaining issue was the compliance schedule. The costs of various lead standards were estimated, using a model of the U.S. refining sector originally developed for the Department of Energy. The model represents the various types of refining capabilities that are available to convert crude oils to final petroleum products. It employs an optimization procedure for finding the allocations of crude oils and intermediate petroleum products among refining units that maximizes social surplus, the sum of consumer and producer surpluses. This allocation corresponds to that which would result from a perfectly competitive market operating without constraints on the utilization of available units. Cost was estimated by looking at the decline in social surplus resulting when the lead constraint was tightened—for instance, from 1.1 gplg to 0.1 gplg. The manufacturers of lead additives challenged these results on the grounds that the model assumed more flexibility in capacity utilization across different refineries than was realistic.

The analytical team held meetings with staffers from other EPA offices to consider alternative compliance schedules. Although a tentative decision was reached to set an interim standard of 0.5 gplg, to be effective July 1, 1985, and a final standard of 0.1 gplg, to be effective January 1, 1986, several staffers feared that some refiners would be unable to comply with their existing equipment. If these fears were realized, the economic costs of the new rule would be higher than estimated and perhaps raise political problems.

A consultant to the project, William Johnson of Sobotka and Company, suggested a solution. If the physical distribution of equipment among refineries interfered with the flexibility in petroleum transfers assumed in the model, he reasoned, perhaps a

[51]V. Batuman, E. Landy, J. K. Maesaka, and R. P. Wedeen, “Contribution of Lead to Hypertension with Renal Impairment,” *New England Journal of Medicine,* 309 (1983), pp. 17–21.

[52]The results of their research were later published in J. L. Pirkle, J. Schwartz, J. R. Landes, and W. R. Harlan, “The Relationship between Blood Lead Levels and Blood Pressure and Its Cardiovascular Risk Implications,” *American Journal of Epidemiology,* 121, 2 (1985), pp. 246–258.

secondary market in lead rights could be created to facilitate trading to get around specific bottlenecks. Taking the total permitted lead content from July 1, 1985 to January 1, 1988 as a constraint, the key was to create an incentive for refiners who could make the least costly reductions in lead additives below the interim 0.5 gplg standard to do so. Their additional reductions could then be used to offset excess lead in gasoline produced by refiners who could not easily meet the basic standards with the equipment they had in place. Because current reductions below the standard create a right to produce above the standard some time in the future, the trading process was called "banking of lead rights." Refiners would be free to buy and sell lead rights at prices that were mutually beneficial. As a result, the aggregate cost of meeting the new standard would be reduced.

Representatives from the various EPA offices involved with the lead rule agreed that banking seemed to be a good way to deal with concerns about the compliance schedule. Because it had not been discussed in the proposed rule published in August, banking could not be part of the final rule. Nevertheless, by moving quickly to propose banking in a supplemental notice, it would be available shortly after the new standard became final.[53]

The remaining task for the team was to prepare the *Final Regulatory Impact Analysis,* which would be published in support of the final rule.[54] The resulting document began by discussing the misfueling and health problems associated with lead additives along with alternatives (public education and stepped-up local enforcement to deal specifically with misfueling, pollution charges to deal generally with lead as a negative externality, and other regulatory standards) to the final rule. It then detailed the methods used to estimate the costs of tighter lead standards, the link between gasoline lead and blood lead, the health benefits of reducing the exposure of children and adults to lead, the benefits of reducing pollutants other than lead, and the benefits from reduced vehicle maintenance costs and increased fuel economy.

The present value of the net benefits of the final rule was presented with various assumptions about misfueling (the use of leaded gasoline in vehicles with catalytic converters). The lower level of permitted lead would reduce the price differential between leaded and unleaded gasoline, thereby reducing the economic incentive for misfueling. It was not possible to predict with confidence, however, how much misfueling would actually decline. Therefore, the reasonable approach was to consider net benefits over the range of possibilities. Table C9.2.2 presents the results of this sensitivity analysis. Note that net benefits were given both including and excluding the adult blood pressure benefits. Although the blood pressure benefits appeared huge, they were the last of the benefit measures considered and hence had the least-developed supporting evidence. Nevertheless, even assuming that the standard would produce no reduction in misfueling and no health benefits for adults, the present value of benefits appeared to be more than double the present value of costs. Indeed, it appeared that maintenance benefits alone would more than cover higher refining costs.

The *Final Regulatory Impact Analysis* was released in February 1985. On March 7, 1985, the

[53]*Fed. Reg.* 718 (January 4, 1985); and 50 Fed. Reg. 13116 (April 2, 1985).

[54]Joel Schwartz, Hugh Pitcher, Ronnie Levin, Bart Ostro, and Albert L. Nichols, *Costs and Benefits of Reducing Lead in Gasoline: Final Regulatory Impact Analysis,* Publication No. EPA-230–05–85–006 (Washington, DC: Office of Policy Analysis, EPA, February, 1985). By the time the final report was prepared, Jane Leggett left the project team. In the meantime, Albert Nichols, a Harvard University professor who was visiting the EPA as the acting director of the Economic Analysis Division, began working closely with the team to produce the final document.

[55]50 *Fed. Reg.* 9386 (March 7, 1985).

[56]See Isaac Ehrlich, "The Deterrent Effect of Capital Punishment: A Question of Life and Death," *American Economic Review,* 65, 3 (June 1975), pp. 397–417; and Alfred Blumstein, Jacqueline Cohen, and Daniel Nagin, eds., *Deterrence and Incapacitation: Estimating the Effects of Criminal Sanctions on Crime Rates* (Washington, DC: National Academy of Sciences, 1978).

[57]See Charles A. Lave, "Speeding, Coordination, and the 55 MPH Limit," *American Economic Review,* 75, 5 (December 1985), pp. 1159–1164; and Peter Asch and David T. Levy, "Does the Minimum Drinking Age Affect Traffic Fatalities?" *Journal of Policy Analysis and Management,* 6, 2 (Winter 1987), pp. 180–192.

[58]See Eric A. Hanushek, "Throwing Money at Schools," *Journal of Policy Analysis and Management,* 1, 1 (Fall 1981), pp. 19–41.

final rule was published in the *Federal Register*.[55] The 0.1 gplg standard would take effect on January 1, 1986, almost three years after work on the supporting analysis began.

Conclusion

We have described a case where statistical analysis made an important contribution to changing policy. Is this case typical? Yes and No. You should not expect that such a confluence of skill, time, data, resources, and interest will often arise to produce such definitive empirical findings. At the same time, you should expect to encounter empirical questions that at least can be approached, if not confidently answered, with the sort of statistical methods used by the EPA analysts. Consider a few prominent examples from recent years: Does the death penalty deter homicide?[56] Do higher minimum legal drinking ages and the 55-mile-per-hour speed limit reduce traffic fatalities?[57] Do smaller class sizes improve student performance?[58] Although widely accepted answers to these empirical questions would not necessarily be decisive in resolving policy debates, they would at least move the debates beyond disputes over predictions to explicit considerations of values. Such highly controversial issues aside, you are likely to find that making empirical inferences and critically consuming those of others often contribute in important ways to the quality of your policy analyses. ■

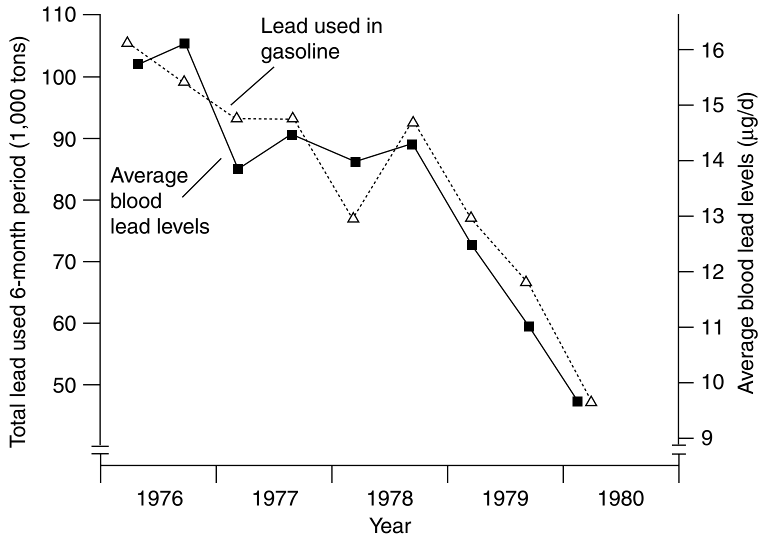


FIGURE C9.2.1
Lead Use in Gasoline Production and Average NHANES II Blood Lead Levels

Source: J. Schwartz et al., *Costs and Benefits of Reducing Lead in Gasoline: Final Regulatory Impact Analysis,* Publication No. EPA-230-05-85-006 (Washington, DC: EPA, 1985), p. E-5.

TABLE C9.2.1

Basic Regression Model for Estimating the Effects of Gasoline Lead on Blood Lead[a]

Effect	Coefficient	Standard Error	*P*-value
Intercept	6.15		
Gasoline	2.14	.142	0.0000
Low income	0.79	.243	0.0025
Moderate income	0.32	.184	0.0897
Child (under 8)	3.47	.354	0.0000
Number of cigarettes	0.08	.012	0.0000
Occupationally exposed	1.74	.251	0.0000
Vitamin C	−0.04	.000	0.0010
Teenager	−0.30	.224	0.1841
Male	0.50	.436	0.2538
Male teenager	1.67	.510	0.0026
Male adult	3.40	.510	0.0000
Small city	−0.91	.292	0.0039
Rural	−1.29	.316	0.0003
Phosphorus	−0.001	.000	0.0009
Drinker	0.67	.173	0.0007
Heavy drinker	1.53	.316	0.0000
Northeast	−1.09	.332	0.0028
South	−1.44	.374	0.0005
Midwest	−1.35	.500	0.0115
Educational level	−0.60	.140	0.0000
Riboflavin	0.188	.071	0.0186
Vitamin A	0.018	.008	0.0355

Note: [a] *Dependent variable:* Blood lead (μg/dl) of whites in NHANES II Survey.

Source: Joel Schwartz et al., *Costs and Benefits of Reducing Lead in Gasoline: Final Regulatory Impact Analysis* (Washington, DC: EPA, 1985), p. III-15. The original document reported incorrect standard errors. The standard errors reported here were provided by Joel Schwartz.

TABLE C9.2.2

Present Values of Costs and Benefits of Final Rule, 1985–1992 (Millions of 1983 Dollars)

	No Misfueling	Full Misfueling	Partial Misfueling
Monetized benefits			
Children's health effects	2,582	2,506	2,546
Adult blood pressure	27,936	26,743	27,462
Conventional pollutants	1,525	0	1,114
Maintenance	4,331	3,634	4,077
Fuel economy	856	643	788
Total monetized benefits	37,231	33,526	35,987
Total refining costs	2,637	2,678	2,619
Net benefits	34,594	30,847	33,368
Net benefits excluding blood pressure	6,658	4,105	5,906

Source: Joel Schwartz et al., *Costs and Benefits of Reducing Lead in Gasoline* (Washington, DC: EPA, 1985), table VIII-8, p. VIII-26.

APPENDIX 1

Policy Issue Papers

The policy issue paper is a report that presents an analysis of a policy issue. The kinds of questions addressed in policy issue papers are these: What actual or potential courses of action are the objects of disagreement among policymakers and other stakeholders? In what different ways may the problem be defined? What is the scope and severity of the problem? If nothing is done, is the problem likely to get worse in future months or years? What goals and objectives should be pursued to solve the problem? How can the degree of success in achieving objectives be measured? What activities are now under way to resolve the problem? What new policy alternatives should be considered? Which are preferable and why?

In answering these questions, the analyst needs to acquire knowledge and skills in developing written materials and communicating them in policy presentations and other briefings. Only recently have such knowledge and skills been included in instructional programs in public policy analysis. Policy analysis may be "applied" or "basic."[1] But the mission of policy analysis is mainly "applied."

FOCUS AND FORMS OF THE POLICY ISSUE PAPER

The policy issue paper may focus on problems at one or more levels of government. Global warming and air pollution, for example, are international, national, and local in scope. The issue paper may take a number of specific forms, depending on the audience and the

[1]The terms "basic" and "applied" are used for convenience only. It is widely acknowledged that these two orientations toward science overlap in practice. Indeed, there is much evidence to support the conclusion that many of the most important theoretical, methodological, and substantive advances in the "basic" social sciences originated in "applied" work on practical problems and conflicts. See, especially, Karl W. Deutsch, Andrea S. Markovits, and John Platt, *Advances in the Social Sciences, 1900–1980*: *What, Who, Where, How?* (Lanham, MD: University Press of America and Abt Books, 1986).

particular issue at hand. Thus, issue papers may be presented in the form of "staff reports," "briefing papers," "options papers," or so-called "white papers." The following is an illustrative list of issues that may serve as the focus of a policy issue paper:

- Which of several alternative contracts should be accepted by a union bargaining team?
- Should the mayor increase expenditures on road maintenance?
- Should the city manager install a computerized management information system?
- Which public transportation plan should the mayor submit for federal funding?
- Should a state agency establish a special office to recruit minorities and women for civil service positions?
- Should a citizens' group support environmental protection legislation now before Congress?
- Should the governor veto a tax bill passed by the state legislature?
- Should an agency director support a plan for flexible working hours (flextime)?
- Should a legislator support a bill restricting the sale of hand guns?
- Should the president withhold foreign aid from countries that violate human rights?
- Should the U.S. Congress repeal the Affordable Care Act?
- Should the United States withdraw from Afghanistan?

ELEMENTS OF THE POLICY ISSUE PAPER

A policy issue paper should be as complete as time and available information permit. An issue paper should "explore the problem at a depth sufficient to give the reader a good idea of its dimensions and the possible scope of the solution, so that it might be possible for a decision-maker to conclude either to do nothing further or to commission a definitive study looking toward some action recommendation."[2] In this author's experience, most issue papers deal primarily with the formulation of a problem and possible solutions. Only rarely does the issue paper reach definitive conclusions or recommendations. Although an issue paper may contain recommendations and outline plans for monitoring and evaluating policy outcomes, it is essentially the first phase of an in-depth policy analysis that may be undertaken at a later time.

In preparing an issue paper, the analyst should be reasonably sure that all major questions have been addressed. Although issue papers will vary with the nature of the problem being investigated, most issue papers contain a number of standard elements.[3] These elements, already presented in Figure 9.2, have been organized around the framework for policy analysis presented in the text.

---

[2]Edgar S. Quade, *Analysis for Public Decisions* (New York: American Elsevier Publishing, 1975), p. 69.

[3]Compare Quade, *Analysis for Public Decisions*, pp. 68–82; and Harry Hatry and others, *Program Analysis for State and Local Governments* (Washington, DC: The Urban Institute, 1976), Appendix B, pp. 139–43.

Element	Policy-Analytic Procedure
Letter of Transmittal	
Executive Summary	
I. Background of the problem	Monitoring
A. Describe client's inquiry	
B. Overview problem situation	
C. Describe prior efforts to solve problem	
II. Significance of the problem	Evaluation
A. Evaluate past policy performance	
B. Assess the scope and severity of problem	
C. Determine the need for analysis	
III. Problem statement	Problem Structuring
A. Diagnose problem	
B. Describe major stakeholders	
C. Define goals and objectives	
IV. Analysis of alternatives	Forecasting
A. Describe alternatives	
B. Forecast consequences of alternatives	
C. Describe any spillovers and externalities	
D. Assess constraints and political feasibility	
IV. Conclusions and recommendations	Prescription
A. Select criteria or decision rules	
B. State conclusions and recommendations	
C. Describe preferred alternative(s)	
D. Outline implementation strategy	
E. Summarize plan for monitoring and evaluation	
F. List limitations and unanticipated consequences	
References	
Appendices	

Observe that each element of the issue paper requires a different policy-analytic procedure to produce and transform information. A policy issue paper, however, has one major characteristic not shared by integrated policy analysis. The issue paper is essentially a prospective (*ex ante*) investigation that begins with limited information about past policy outcomes and performance and ends with as much information as possible about the nature of policy problems, expected policy outcomes, and preferred policies.

A CHECKLIST

This checklist is designed as a practical guide to the preparation and self-evaluation of policy issue papers. Items in the checklist are based on guidelines for preparing policy issue papers presented earlier in this appendix. The checklist operationalizes these guidelines, provides brief examples, and contains a rating scale that may be used to evaluate the adequacy of specific elements of the paper. Remember, however, that some elements in the checklist may not be appropriate for each and every paper. The checklist should therefore be used in a flexible and creative manner.

Rating Scale

1 = totally adequate
2 = adequate
3 = inadequate
4 = totally inadequate
0 = not applicable

Element of Policy Issue Paper	Example/Note	Rating
Letter of transmittal		
1. Are all relevant stakeholders expected to act on the paper addressed?	A paper on local law enforcement might be addressed to the chief of police, the mayor, and the director of public safety.	[]
2. Does the letter describe all attached materials?	"Enclosed please find a policy issue paper on options to increase the productivity of municipal employees. An executive summary of this paper is also included, along with relevant statistical materials appended to the text."	[]
3. Does the letter specify who is expected to take action, how, and when?	"We look forward to your reply well in advance of the meeting scheduled for June 15."	[]
Executive summary		
4. Are all elements of the issue paper described in the executive summary?	Every element of the issue paper (source and background of the problem, the policy problem, and so on) should be included in the summary.	[]
5. Is the summary clear, concise, and specific?	"This paper reviews problems of air pollution over the past 10 years, showing that the level of industrial pollutants has increased by more than 200 percent in the 1968–1978 period."	[]
6. Is the summary understandable to all who will read it?	The point here is to avoid unnecessary jargon, long sentences, and complex arguments as much as possible.	[]

Element of Policy Issue Paper	Example/Note	Rating
7. Are recommendations appropriately highlighted in the summary?	"In conclusion, we recommend that funds in the range of $15,000–20,000 be allocated to conduct an assessment of recreational needs in the Oakland area."	[]
I. Background of the problem		
8. Are all dimensions of the problem situation described?	Increasing crime rates may be analyzed in conjunction with data on unemployment, police recruitment, citizen surveys, migration, and so on.	[]
9. Have outcomes of prior efforts to resolve problems in the area been described?	"Similar training programs have been implemented in New York, Florida, and California, at a cost of $4,000–5,000 per trainee."	[]
10. Is there a clear assessment of past policy performance?	"Training programs in New York and California resulted in a 1 percent reduction in hard-core unemployment and were judged to be highly successful."	[]
II. Scope and severity of problem		
11. Is the scope and severity of the problem situation clearly described?	"Problems of industrial pollution are more severe today than at any time in our history and affect the physical well-being of 60 percent of the population. The resolution of these problems is an urgent matter that calls for timely and decisive action."	[]
III. Problem statement		
12. Is the problem clearly stated?	"The problem addressed in this paper is how best to satisfy the need for increased fiscal accountability among local governments in the region."	[]
13. Is the issue clearly stated?	"The issue addressed in this paper is whether the state's Department of Community Affairs should increase its technical assistance to local governments in the area of financial management."	[]
14. Is the approach to analysis clearly specified?	"In addressing this issue we have employed cost–effectiveness analysis as a way to determine the levels of investment necessary to train 10,000 local officials in cash management techniques."	[]

(*continued*)

Element of Policy Issue Paper	Example/Note	Rating
15. Are all major stakeholders identified and prioritized?	The task here is to identify all persons and groups who significantly affect and are significantly affected by the formulation and implementation of policies. Stakeholders who are only marginally influential (or affected) may be dropped from the analysis, but this requires prioritization.	[]
16. Are goals and objectives clearly specified?	"The goal of this policy is to improve the physical security of the community. The objective is to reduce reported crimes by 10 percent or more in the 1979–1980 period."	[]
17. Are measures of effectiveness clearly specified?	Here there are many choices: benefit–cost ratios, net benefits, distributional benefits, and so on.	[]
18. Are all sets of potential solutions outlined?	"Any effort to reduce crime in the area should take into account the effectiveness of law enforcement programs, the unemployment rate, population density, and urbanization."	[]
IV. Policy alternatives		
19. Are alternative solutions specified?	"Three policy alternatives are analyzed in the course of this paper: educational programs to alert citizens to the role they can play in crime control; policy training programs; and advanced crime control technology."	[]
20. Are alternatives systematically compared in terms of their probable costs and effectiveness?	"Program A has a benefit–cost ratio that is twice that of Program B."	[]
21. Are relevant spillovers and externalities included in the analysis of alternatives?	"The early childhood educational program is not only cost-effective. It will also produce spillovers in the form of increased participation of family members in the school. At the same time there are important externalities in the form of increased costs for transportation to and from school."	[]
22. Are all relevant constraints taken into account?	These may be financial (e.g., fixed costs); legal (e.g., laws proscribing certain actions); or political (e.g., lack of political support).	[]
23. Have alternatives been systematically compared in terms of political feasibility?	The point here is to make informed judgments about the probable support, relevance, and influence of key stakeholders in gaining acceptance for and implementing each alternative.	[]

Element of Policy Issue Paper	Example/Note	Rating
V. Policy recommendations		
24. Are all relevant criteria for prescribing alternatives clearly specified?	Criteria may include: net welfare improvement, distributional improvement, cost effectiveness, and so on.	[]
25. Has the preferred alternative been clearly described?	"We therefore recommend that the second alternative be adopted and implemented, since it is likely to be more cost-effective and is politically feasible under present conditions."	[]
26. Is a strategy for implementation outlined?	"The proposed policy should be implemented by the Department of Labor and state employment offices. Implementation should be phased over a period of 3 years and focus on cities with more than 50,000 persons."	[]
27. Are provisions made for monitoring and evaluating policies?	"The state employment offices should establish special units to engage in process and outcome evaluations, at intervals of 30 days. Results of monitoring and evaluation should be stored in a management information system and reports should be filed quarterly."	[]
28. Are limitations and possible unintended consequences taken into account?	Here the point is to specify how limitations of the analysis affect the confidence one has in the conclusions. Possible unintended consequences (e.g., abuses of social welfare programs, nuclear meltdowns, and so on) should also be thought through.	[]
References		
29. Are references cited in the text included in the appropriate form?	It is preferable to use the "name–date" bibliographic form in a policy issue paper. Only the name of the author, the date of publication, and page numbers will be included in the text, with no footnotes at the bottom of the page. For example: (Jones, 1975: 20–23). The citation in the references will read: Jones, John (1975), *A Report on the Fiscal Crisis*. Washington, DC: Fiscal Institute.	[]
Appendices		
30. Is relevant supporting information (statistics, legislation, documents, correspondence) included?	All statistical tables should be properly numbered and headed and the source of data specified at the bottom of the table. Legislation, documents, and correspondence should be identified as to its source, date, author(s), and so on.	[]

APPENDIX 2

Executive Summaries

The executive summary is a synopsis of a policy issue paper. The executive summary usually has these elements:

- Purpose of the issue paper or study being summarized
- Background of the problem or question addressed
- Major findings or conclusions
- Approach to analysis or methodology
- Recommendations (optional: depends on expectations of the client)

The following executive summary provides a synopsis of a 90-page study titled *Freight Trucking: Promising Approach for Predicting Carriers' Safety Risks* (Washington, DC: U.S. General Accounting Office, Program Evaluation and Methodology Division, April 1991).

	Executive Summary
Purpose	Freight trucks pose special safety risks. Over 4,000 people are killed annually in accidents related to heavy trucks. Fatalities are about twice as likely in accidents involving tractor-trailer trucks as in those involving automobiles only. In recent years, the Congress has approved legislation to prevent situations that give rise to unsafe trucking operations. As a means toward this end, the House Committee on Public Works and Transportation and its Surface Transportation Subcommittee requested that United States General Accounting Office (GAO) determine whether certain economic and other conditions could be used as predictors of safety outcomes. GAO's study had the following three objectives: (1) to formulate a predictive model specifying hypothetical relationships between safety and a set of conditions in the trucking industry; (2) to assess the availability and quality of federal data required to

test the model; and (3) to use available data, to the extent possible, to develop a set of indicators that would predict safety problems in the freight-trucking industry. The value of a workable model is that the Department of Transportation (DOT) could use it as an early warning system for predicting safety problems.

Background

Although the Motor Carrier Act of 1980 codified the relaxation of federal economic control over the trucking industry, the Congress approved legislation in the 1980s designed to monitor and prevent situations that result in unsafe trucking operations.

GAO developed a model that hypothetically links changes in economic conditions to declining safety performance in the freight-trucking industry (see pages 18 through 23). The hypothesis is that a decline in economic performance among motor carriers will lead to declining safety performance in one or more ways, described by five submodels: (1) a lowering of the average quality of driver performance; (2) downward wage pressures encouraging noncompliance by drivers with safety regulations; (3) less management emphasis on safety practices; (4) deferred truck maintenance and replacement; and/or (5) introduction of larger, heavier, multitrailer trucks.

Results in brief

GAO's preliminary findings, using data on 537 carriers drawn from both DOT and the Interstate Commerce Commission (ICC), are that seven financial ratios show promise as predictors of safety problems in the interstate trucking industry. For example, three measures of profitability—return on equity, operating ratio, and net profit margin—were associated with subsequent safety problems as measured by accident rates. The data agreed with GAO's model for five of seven financial ratios: Firms in the weakest financial position had the highest subsequent accident rates. GAO also used a number of other factors to predict safety outcomes, including the following. First, the smallest carriers, as a group, had an accident rate that exceeded the total group's rate by 20 percent. Second, firms operating closer to a broker model—that is, those that rely on leased equipment and/or drivers to move freight—had a group accident rate 15–21 percent above the total group's rate.

With regard to two of the submodels (driver quality and compliance), driver's age, years of experience, and compensation were all good predictors of safety problems. GAO's evidence is generally consistent with the model's hypotheses because younger, less experienced drivers and lower paid company drivers posed greater-than-average accident risks.

GAO's study thus demonstrates the potential for developing preventive strategies geared to differences among carriers and drivers, and it also suggests the importance of monitoring by DOT of the variations in carrier accident rates in order to have a sound basis for developing those preventive strategies.

GAO's Analysis

Available Federal Data

To identify and evaluate data to test a carrier-safety model, GAO reviewed the literature, talked with industry experts, and conducted interviews with federal officials responsible for maintaining data sets. GAO then combined data provided by DOT and ICC to conduct analyses. GAO found that existing federal data sets did not

(*continued*)

bring together the necessary data to fully test this model. The federal collection of truck accident data was essentially independent of the gathering of economic data, and combining the two types of data from separate federal sources was generally impractical. Most importantly, the federal data allowing calculation of accident rates for individual motor carriers did not provide for a generalizable picture of a definable segment of the industry or an analysis of safety trends over time. The needed information about truck drivers and their accident rates was also lacking. As a result, GAO could test only two of the submodels (by obtaining data from two private surveys). One unfortunate implication of this is that even if all of the submodels do prove to have predictive validity, existing federal databases still do not contain sufficient information to convert the model to an effective monitoring system.

Economic Predictors

GAO judged that the best available accident rate data to combine with ICC's extensive financial data are those obtained from DOT's safety audits. Since the safety audits were discontinued after October 1986, GAO's analysis was limited to the larger, for-hire ICC carriers with financial reporting requirements that were also audited by DOT during the years 1984–1986.

GAO found evidence among these interstate carriers that carriers in different markets or different financial situations pose different safety risks. For example, carriers with losses of 0.3 percent or more on equity had a group accident rate (rates are defined as accidents per million miles) 2 years later that was 27 percent above the overall group's rate.

Predictors from the Driver Quality Submodel

One of the private surveys GAO used supplied data on approximately 1,300 interstate drivers serving Florida in 1989. As was predicted by the driver quality submodel, GAO found that younger and less experienced truck drivers were more likely to be in accidents. For example, the odds for drivers aged 21 to 39 having been involved in an accident in the prior 12 months were higher than the odds for drivers over age 49 by a factor of 1.6.

Predictors from the Driver Compliance Submodel

The other private survey GAO used yielded pertinent data from a national sample of drivers in rail-competitive trucking. GAO found that lower paid drivers were more likely than their higher paid counterparts to violate safety regulations, but only in the case of company drivers and excluding owner-operators (those drivers owning their own trucks). Among company drivers, those earning less than 18.5 cents per mile had about twice the odds of having received either speeding or hours-of-service citations (or warnings) in the past 90 days.

Recommendation to the Secretary of Transportation

The monitoring, enforcement, and policymaking value of much of the truck accident information gathered by DOT is lessened by the inability to construct accident rates. Although DOT already collects accident *data*, the mileage data required to calculate accident *rates* are not routinely collected from carriers. As a first step toward reducing the accidents of motor carriers, GAO therefore recommends that the Secretary of Transportation direct the Administrator of the Federal Highway Administration (FHWA) to require that mileage data on motor carriers falling under FHWA regulations be obtained annually to improve accident analysis. How such data are obtained may depend on a number of considerations, such as costs and respondent burden, but the foremost consideration should be that data obtained allow for the calculation of accident rates for carriers falling under FHWA safety regulations in order to support monitoring and enforcement efforts and to permit analysis of safety trends.

In implementing GAO's recommendation, DOT should consider further development of predictors of safety problems. For example, GAO's analysis suggests that indicators of financial health, market segment, and driver information may be useful to DOT in identifying higher risk groups of carriers for closer monitoring or enforcement efforts. More work needs to be done in validating these preventive indicators and identifying other predictors of safety outcomes. DOT should consider advancing this work on preventive indicators because, if successful, it would signal the policy changes needed to avoid or abate the predicted unsafe conditions. GAO's demonstration illustrates the kind of work that DOT will be able to do in prevention, particularly if better information on accident rates and economic and other intermediate factors is developed.

APPENDIX 3

Policy Memoranda

Policy issue papers are the product of a lengthy process requiring many months or years of research and analysis. Policy memoranda, by contrast, are prepared over a much shorter period of time—often no more than one month, but sometimes a matter of weeks or days. Policy memoranda are often based on syntheses of one or many issue papers, studies, or reports. The differences between the issue paper, on one hand, and the policy memorandum on the other are illustrated by contrasts between "basic" and "applied" policy analysis.

TWO KINDS OF POLICY ANALYSIS

Characteristic	Basic	Applied
Origin of problems	• University community	• Policy community
Typical methods	• Quantitative modeling	• Sound argumentation
Type of research	• Original data collection	• Synthesis of existing data
Primary aim	• Improve theory	• Improve practice
Communications media	• Article or book	• Policy memo or issue paper
Source of incentives	• University departments	• Government departments

The policy memo (short for *memorandum*) is a form of written communication used in public, nonprofit, and business organizations. Policy memos are brief, well organized, and understandable to non-specialists. In some cases, they include policy recommendations (but only if requested). The clients for policy memos may be appointed elected officials, citizens groups, or representatives of nonprofit organizations, nongovernmental organizations, and business organizations.

Policy memos take different forms, depending on the agency or organization. For example, different agencies, departments, or ministries have their own letterhead stationery that is preformatted, while others use word-processing templates. Despite these differences, there are several common elements. Most memos display the date, the name of the recipient, the name of the analyst, and the subject of the memo on separate lines at the top of the page. Because memos are designed to communicate with busy policymakers, memos should be brief (most are no more than three to five pages in length), clearly organized with appropriate headings, and understandably written with as little technical jargon as possible. The content of the memo should be concisely communicated on the subject line. Technical and analytic material (tables, graphs, charts, legal documents, etc.) that supports the conclusions and recommendations in the memo should be appended to the memo as an attachment, or kept on hand to respond to questions that may be (and often are) raised later.

- The introductory paragraph should state the purpose of the memo and the question or issue to which the memo is a response.
- There may be a brief description of the methods used in doing the analysis. But remember that busy readers usually do not care very much about your methods. (This does *not* mean that you will not have to justify your methods to someone else later.)
- As appropriate, longer memos with about three pages of more should use headers, bullets, underlining, italics, and other stylistic devices to highlight material that is especially important.
- Normally, the memo should end with information about where you can be reached for further questions (e.g., telephone, fax, e-mail).

Following are examples of a cross-section of policy memos.

POLICY MEMO

TO: The President-Elect's National Security Transition
FROM: David Shorr, The Stanley Foundation
Vikram Singh, Center for a New American Security
RE: A Unified International Affairs and National Security Budget to Increase American Effectiveness Worldwide

The Center for a New American Security (CNAS) and Stanley Foundation recently concluded an eight-month initiative focused on the anemic condition of the United States' civilian international affairs agencies. There is nearly universal support among national security professionals for serious action to remedy this condition. Defense Secretary Gates continues to call for more resources for the civilian elements of national power and the Project on National Security Reform is expected to recommend a doubling of State Department and USAID funding and a minimum manpower increase of 50%. Yet this widely recognized need for stronger civilian agencies has not been paired with the political will to mandate and underwrite them. 94% of supplemental funding for the wars in Iraq and Afghanistan has been for the Department of Defense.

The essence of the problem is inadequate financial and human resources to sustain US political influence in a fast-changing world. Just when US relations with the rest of the world are in a deep slump, our capacity to turn things around remains at historic lows.

America needs foreign policy infrastructure investment, and not just for special initiatives or boutique programs.

The complete findings of the CNAS – Stanley project are in four reports available at http://www.stanleyfoundation.org/articles.cfm?ID=492. This memo, however, proposes a single major policy initiative to give immediate impetus and lay the foundation for the rebuilding of America's international affairs capabilities. *The President-elect should require his new national security and budget teams to prepare a joint FY2010 international affairs and national security budget.*

The Obama Administration will, and should, consider specific measures to redress organizational weaknesses in the US government's foreign policy apparatus, particularly post-conflict reconstruction, long-term economic development, and public diplomacy. It is critical to recognize, though, that these are all symptoms of a systemic civilian capacity deficit. Since this is a matter of the overall effectiveness of US foreign policy, the push for budgeting across interagency stovepipes must come from the top. In other words, whatever other reforms are undertaken, a presidential mandate for a joint FY2010 budget is vitally important to initiate the process of rebuilding organizational infrastructure.

The joint budget for international affairs and national security—starting with the combining of the 050 and 150 budget functions for FY2010 and expanding in subsequent years—would at first be a tool to force cross-department balancing. Without mandating changes in normal budget processes within agencies, an integrated White House review will promote collaboration across overlapping budget areas. It will drive:

- Substantial investment in strengthening the State Department and USAID, to bolster their overstretched headquarters bureaus and overseas missions.
- Significant growth in their work forces.
- Steady balancing of the relative increases in defense and international affairs spending seen in recent years.

Detailed budget proposals and implementation plans are available from other initiatives and organizations, such as the *Foreign Affairs Budget for the Future* report from the American Academy of Diplomacy and the Stimson Center. We suggest that a combined international affairs and national security budget is the only lever that can spur more-than-marginal added investment.

This initiative should not be contingent upon, nor await, other reforms to the foreign policy apparatus. A number of related mechanisms will be needed to implement this budgeting discipline, and we urge that they be incorporated into the structures that the new administration is establishing for national security planning and decision-making:

- As these joint budgets are prepared, presented, and enacted, they will need substantial White House staff support from the NSC and OMB. Implementing such a joint budget will require a single dual-hatted NSC and OMB Director for international affairs and national security and integrated reviews of all agency budget submissions.
- Intensive cooperation will be needed among DoD, DoS, and USAID budget offices to help their principals fulfill the president's mandate.
- The budget must be treated as a single package from the beginning to the end of the process. The principals on the president's national security team must hew to shared positions all the way through to the final appropriations bills.

- While reforms of the congressional budget process may ultimately be needed, the administration should meanwhile consult as widely as possible with the budget, authorizing, and appropriating committees. Congress will likely split up such a combined budget upon receipt, but the cross-leveling achieved by this new White House process will still be reflected in the budget bills for the agencies.

Despite being a global power, the United States does not really seem to have its finger on the global pulse. In an era of stakeholder proliferation, every international problem confronting the United States involves not just national governments but a range of actors from the private sector and powerful NGOs to criminal and terrorist networks. To have any chance of shaping world events, Washington must be able to engage this panoply of actors around the globe.

Having drawn down US civilian capabilities faster than defense in the 1990s and increased them more slowly than defense since 2001, the American government has crippled its ability to have insight into, and relationships with, the actors on whom future peace and prosperity hinge. Just as the United States needs to invest in education and science to ensure that the American work force can compete and thrive in the globalizing world, it must likewise transform our government to be competitive in the effort to sustain America's global power in the twenty-first century.

Source: www.stanleyfoundation.org/articles.

August 28, 1986

The Honorable Butler Derrick
Chairman, Task Force on the Budget Process
Committee on the Budget
House of Representatives

Dear Mr. Chairman:

In response to your letter of May 8, 1985, we have examined the potential impacts of immigration on the federal budget over the next 10 years. This effort was undertaken with dual objectives: obtaining the best possible estimate of budgetary impacts despite known data inadequacies and beginning the development of a methodology for dealing with long-term, crosscutting issues of budgetary significance, of which immigration is a prime example. On May 1, 1986, we presented a briefing to you and other task force members and staff on the results of our study. This report is a written version of that briefing. We have also included additional detail on the budgetary impact of immigration on state and local governments.

Our study established fiscal year 1984 as a baseline. We identified, and where necessary estimated, major federal immigrant-related outlays and revenues for that period. We then projected these data to yield 1990 and 1995 estimates under three different sets of assumptions (scenarios) as to the future social, economic, and political environment. These projections are not intended as forecasts but were developed to provide illustrative ranges. We also sought similar data for selected programs and localities in the five most impacted states, for illustrative purposes, but we made no projections at this level.

Although there are constraints and limitations to our analysis, it demonstrates that significant budgetary impacts of immigration exist at the federal level, in terms of both outlays and revenues. Dollar amounts can be assigned to such impacts only if major assumptions are made concerning immigrant participation in major social and income security programs. This is due to the uncertainties of population size and characteristics in 1984, as well as to the lack of direct information regarding immigrant-related outlays and revenues, data gaps which are even more apparent in the out year projections. We assumed, therefore, that these participation rates would be comparable, on the average, to those of U.S. citizens. We could not test this hypothesis, and expert opinion varies.

Using these assumptions, we found that for fiscal year 1984 in the programs we examined

—per capita federal outlays for immigrants were roughly comparable to those for average U.S. residents;
—per capita, immigrants contributed fewer revenues to the federal government than average U.S. residents; and
—total immigrant-related outlays differed only slightly from total immigrant-related revenues.

Under all three scenarios examined, these observations remain valid for 1990 and 1995, although uncertainty increases.

Regarding the broader methodological interests which it was designed to explore, this study of a crosscutting issue has both value and limitations. While specific total dollar amounts cannot be assigned on any supportable basis, minimum outlays and revenues can be identified, ranges can be estimated, and comparisons can be drawn based on explicit assumptions. Because both federal outlays and revenues are significantly affected by immigration and remain roughly in balance, the committee may wish to monitor the key conditions and trends to determine whether this balance is being maintained. Thus, we believe we have demonstrated the feasibility of this approach, but its usefulness will vary with the circumstances.

I would be pleased to discuss this information with you further at your convenience. As agreed with your office, we have not obtained agency comments on this report. We are sending copies of this report to interested parties, and copies will be made available to others upon request.

If you or your staff have any questions on this report, please call me at 275–9455.

Sincerely yours,

Kenneth W. Hunter
Assistant to the Director

Source: *Budget Issues: Immigration to the United States—Federal Budget Impacts, 1984–1995* (Washington, DC: U.S. General Accounting Office, August 1986).

Another type of memo is the briefing memo which summarizes the results of an intelligence analysis. Following is the briefing memo on Osama Bin Ladin's plans for the attack on the World Trade Center on September 11, 2002. This briefing memo was presented to President George W. Bush on August 6, 2001.

DECLASSIFIED PRESIDENTIAL DAILY BRIEFING FOR AUG 6, 2001

For the President Only
6 August 2001

BIN LADIN DETERMINED TO STRIKE IN US

Clandestine, foreign government, and media reports indicate Bin Ladin since 1997 has wanted to conduct terrorist attacks in the US. Bin Ladin implied in US television interviews in 1997 and 1998 that his followers would follow the example of World Trade Center bomber Ramzi Yousef and "bring the fighting to America."

After US missile strikes on his base in Afghanistan in 1998, Bin Ladin told followers he wanted to retaliate in Washington, according to a (redacted) service.

An Egyptian Islamic Jihad (EIJ) operative told a (redacted) service at the same time that Bin Ladin was planning to exploit the operative's access to the US to mount a terrorist strike.

The millennium plotting in Canada in 1999 may have been part of Bin Ladin's first serious attempt to implement a terrorist strike in the US. Convicted plotter Ahmed Ressam has told the FBI that he conceived the idea to attack Los Angeles International Airport himself, but that Bin Ladin lieutenant Abu Zubaydah encouraged him and helped facilitate the operation. Ressam also said that in 1998 Abu Zubaydah was planning his own US attack.

Ressam says Bin Ladin was aware of the Los Angeles operation.

Although Bin Ladin has not succeeded, his attacks against the US Embassies in Kenya and Tanzania in 1998 demonstrate that he prepares operations years in advance and is not deterred by setbacks. Bin Ladin associates surveilled our Embassies in Nairobi and Dar es Salaam as early as 1993, and some members of the Nairobi cell planning the bombings were arrested and deported in 1997.

Al-Qa'ida members—including some who are US citizens—have resided in or traveled to the US for years, and the group apparently maintains a support structure that could aid attacks. Two al-Qa'ida members found guilty in the conspiracy to bomb our Embassies in East Africa were US citizens, and a senior EIJ member lived in California in the mid-1990s.

A clandestine source said in 1998 that a Bin Ladin cell in New York was recruiting Muslim-American youth for attacks.

We have not been able to corroborate some of the more sensational threat reporting, such as that from a (redacted) service in 1998 saying that Bin Ladin wanted to hijack a US aircraft to gain the release of "Blind Shaykh" 'Umar 'Abd al-Rahman and other US-held extremists.

Nevertheless, FBI information since that time indicates patterns of suspicious activity in this country consistent with preparations for hijackings or other types of attacks, including recent surveillance of federal buildings in New York.

The FBI is conducting approximately 70 full field investigations throughout the US that it considers Bin Ladin-related. CIA and the FBI are investigating a call to our Embassy in the UAE in May saying that a group of Bin Ladin supporters was in the US planning attacks with explosives.

Source: www.cnn.com/2004/images/04/10/whitehouse.pdf

APPENDIX 4

Planning Oral Briefings

Written documents are only one means for communicating the results of policy analysis. The other is the oral briefing. Although the oral briefing or presentation has most of the same elements as the issue paper, the approach to planning and delivering an oral presentation is quite different.

The elements of an oral briefing typically include the following:

- Opening and greeting to participants
- Background of the briefing
- Major findings of the analysis
- Approach and methods
- Data used as basis for the analysis
- Recommendations
- Questions from participants
- Closing

As we saw in Chapter 9, an essential aspect of oral communications is knowing the audience. Here a series of important questions should be asked prior to the briefing. How large is the audience? How many members of the audience are experts on the problem you are addressing in the briefing? What percent of the audience understands your research methods? Are you credible to members of the audience? Does the audience prefer detailed or general information? Where does the briefing fall in the policymaking process? At the agenda-setting or formulation and adoption phases? Or the implementation and evaluation phases?

Answers to these questions govern the choice of a communications strategy appropriate for a particular audience. For example, if the audience is a medium-sized group with twenty to twenty-five persons, the diversity of the audience calls for these specific communications strategies.[1]

- Circulate background materials before the meeting so that everyone will start from a common base.
- Tell the group that the strategy selected is designed to meet the objectives of the group as a whole but that it may not satisfy everyone's needs.
- Focus on the agendas of key policymakers, even if you lose the rest of the audience.

Another common problem in giving oral briefings is that policymakers may bring their own staff experts to the briefing. In this case, a high percentage of the audience has expertise in the area you are addressing. Here, the following strategies are appropriate:

- Avoid engaging in a "knowledge contest" with experts by focusing on the purpose of your own presentation.
- Capitalize on the group's expertise by focusing on evidence bearing on findings and recommendations.
- Avoid a long presentation of background material and lengthy description of methods.
- Attempt to generate dialogue and a productive debate on potential solutions for the problem.

An equally important problem in oral briefings is understanding individual members of the group.[2] For example, it is important to know if the immediate clients for policy analysis are:

- Experts in the problem area
- Familiar with your analytic methods
- Influential participants in the policymaking process
- Oriented toward detailed or broad information
- Faced with high, medium, or low stakes in the problem
- Politically important to the success of your recommendations

Finally, oral briefings employ various kinds of graphic displays.[3] One of these is the transparency sheet used with an overhead projector (also known as a "beamer"). In using overhead transparencies, it is important to recognize that they are not appropriate if you want an informal style of presentation—overheads tend to make briefings formal, making open dialogue difficult. Alternatively, you may want to limit discussion, and strict adherence to transparencies helps.

[1]The following guidelines are based on *Presentation Planner* (Pitsford, NY: Discus Electronic Training, Eastman Technology, 1991).
[2]*Presentation Planner*, Individual Questions Menu.
[3]*Presentation Planner*, Graphics Submenu.

It is also important to observe the following guidelines in using transparencies:

- Keep transparencies simple, with condensed and highlighted text.
- Use clear, bold, uncluttered letters and lines of text.
- Highlight important points and use a few different colors (but no more than three).
- Limit the text on any page to a maximum of ten lines.
- Be sure the width of the screen is at least one-sixth of the distance between the screen and the farthest viewer.
- Make sure that the nearest viewer is a minimum of two screen widths from the screen.
- Remain on the same overhead for a minimum of 2 minutes to give viewers time to read plus an additional 30 seconds to study the content.
- Leave the room lights *on*—overheads are not films.
- Turn off the overhead projector when you want to redirect attention back to you and reestablish eye contact.

These are but a few of the strategies and tactics that are appropriate for different audiences and communications media. Given the goal of selecting communications media and products that match the characteristics of the audience, there is today a rich body of theory and techniques to guide the visual display of information.[4] Appropriate computer graphics, presentation packages and statistical programs are now widely available. These include the *Statistical Package for the Social Sciences (SPSS), Microsoft Excel,* and *Microsoft Power Point*.

Many types of visual displays are available to the analyst who wishes to effectively communicate complex ideas. Examples of the following visual displays include the following:

- Bar charts
- Pie charts
- Frequency histograms
- Relative frequency histograms
- Ogives
- Scatterplots
- Time-series graphs
- Graphs with data tables
- Decision trees and other diagrams
- Arrow diagrams and causal models

These displays are useful for effectively communicating simple as well as complex aspects of policies. In most policy presentations, however, a key visual display is one that summarizes relationships between two or more alternatives and their respective impacts. The options–impacts matrix (see Figure A.4.1), which is a rectangular array of policy alternatives and their anticipated impacts, is well suited for this purpose. The matrix provides

[4]Anyone who wishes to achieve competency in addressing these questions should read Edward R. Tufte's award-winning trilogy: *The Visual Display of Quantitative Information* (Cheshire, CT: Graphics Press, 1983); *Envisioning Information* (Cheshire, CT: Graphics Press, 1990); and *Visual Explanations: Images and Quantities, Evidence and Narrative* (Cheshire, CT: Graphics Press, 1997).

a concise visual map of the range of impacts associated with alternatives. The preferred impacts may be shaded, as in Figure A.4.1, so that participants in a presentation can easily examine the pattern of tradeoffs between impacts before they have been represented quantitatively in the form of coefficients, indexes, or ratios (e.g., a benefit–cost ratio), which can obscure issues and prevent informed discussion.

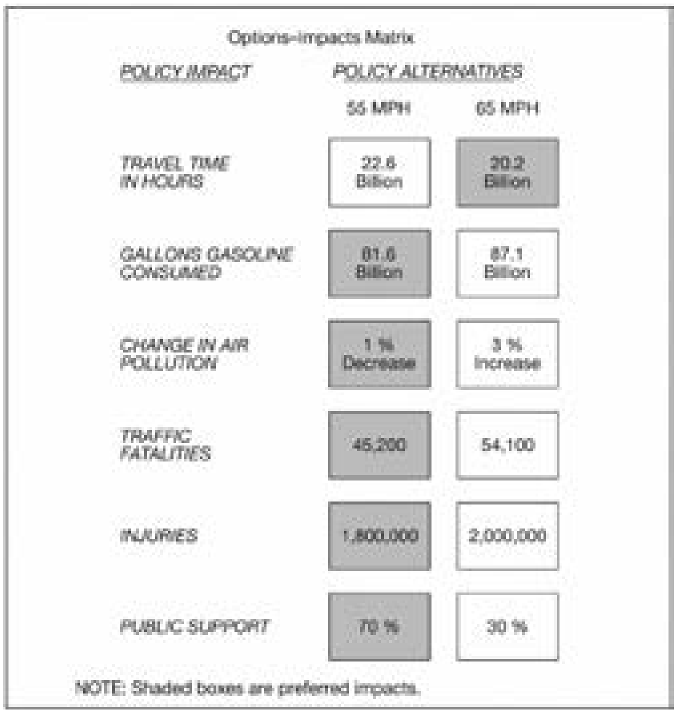


FIGURE A.4.1
Options–Impacts Matrix

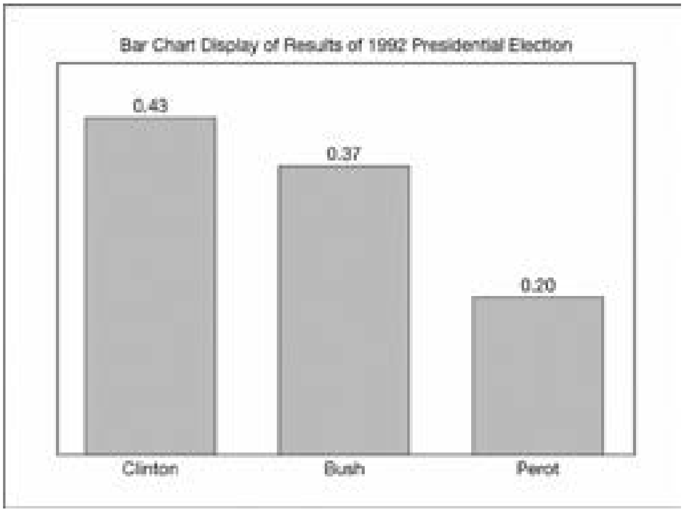


FIGURE A.4.2
Bar Chart Display of Results of 1992 Presidential Election

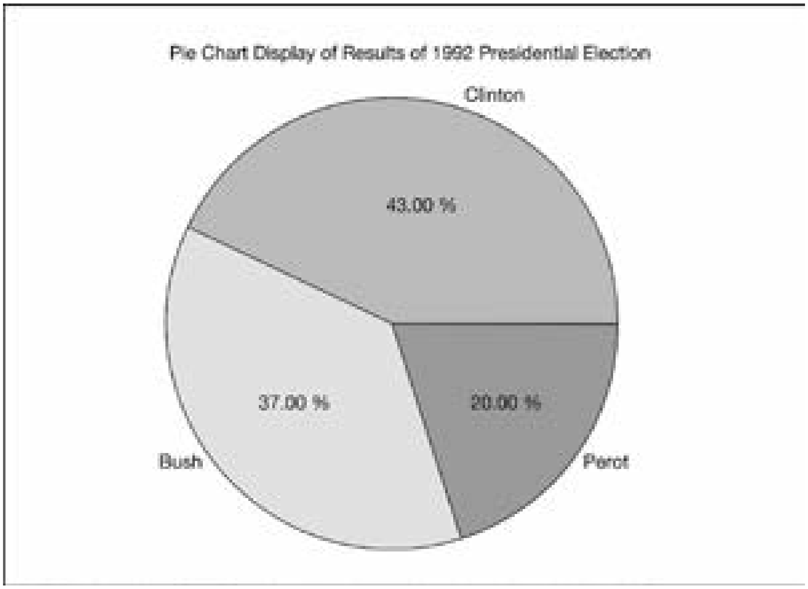


FIGURE A.4.3
Pie Chart Display of Results of 1992 Presidential Election

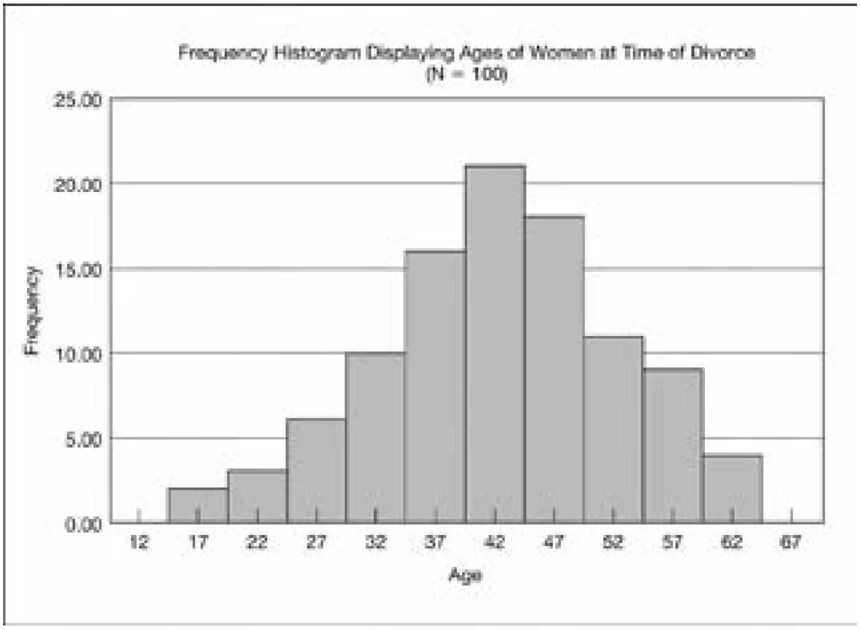


FIGURE A.4.4
Frequency Histogram Displaying Ages of Women at Time of Divorce (N = 100)

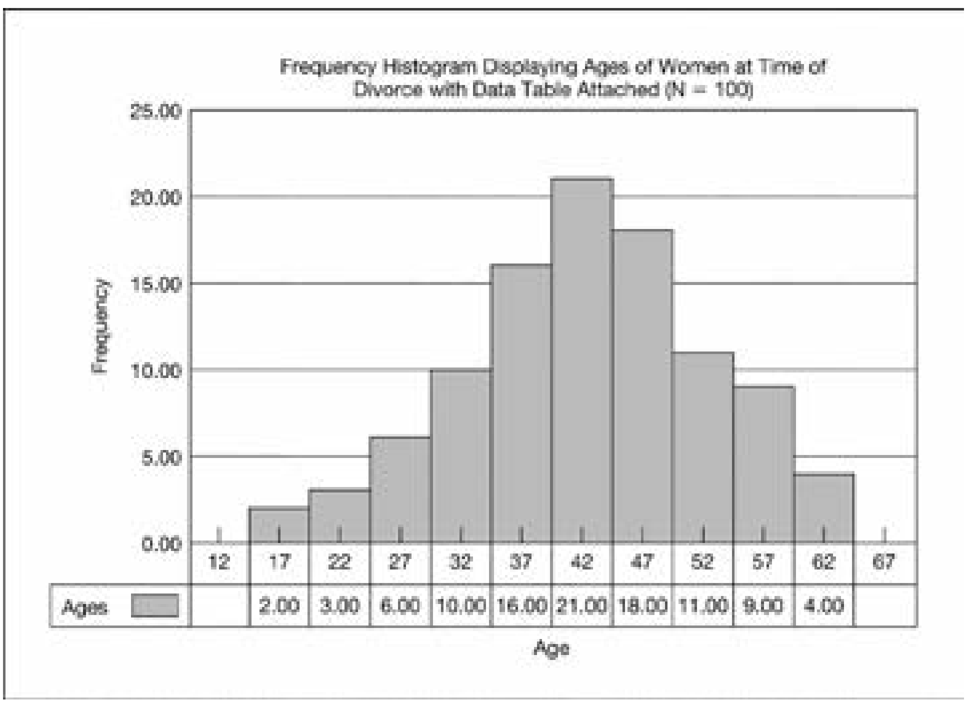


FIGURE A.4.5
Frequency Histogram Displaying Ages of Women at Time of Divorce with Data Table Attached (N = 100)

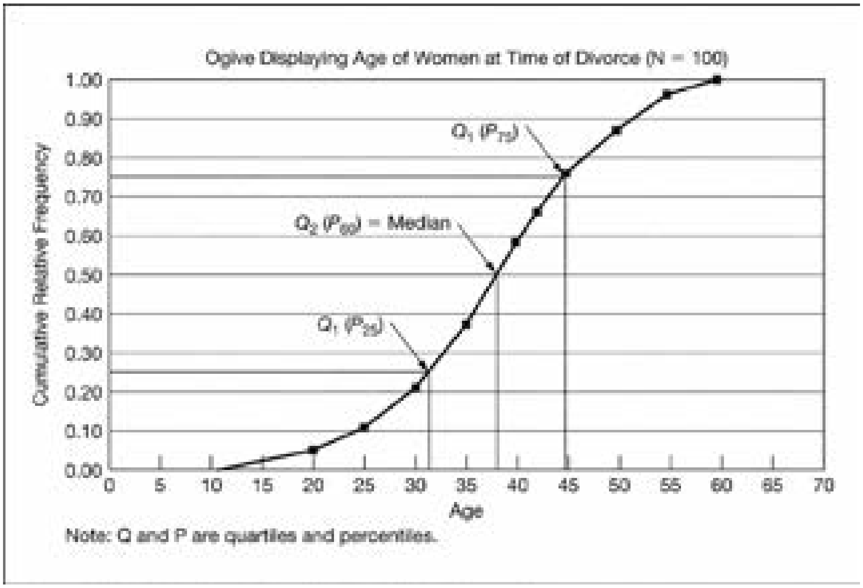


FIGURE A.4.6
Ogive Displaying Ages of Women at Time of Divorce (*N* = 100)

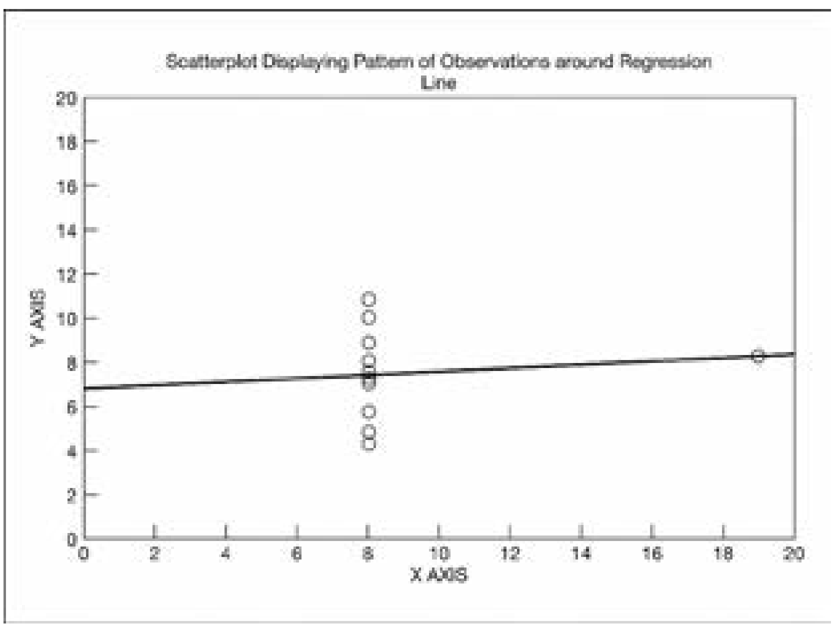


FIGURE A.4.7
Scatterplot Displaying Pattern of Observations around Regression Line

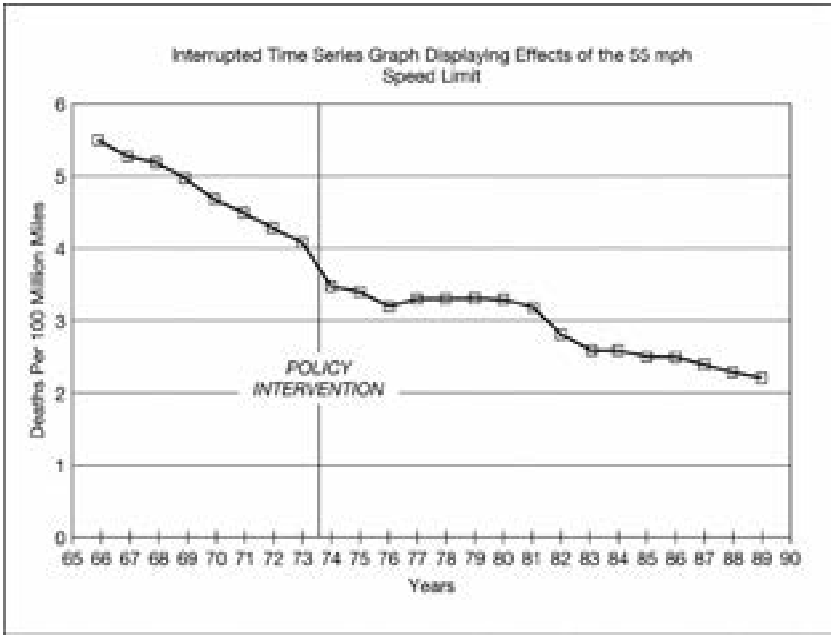


FIGURE A.4.8
Interrupted Time-Series Graph Displaying Effects of the 55 mph Speed Limit

AUTHOR INDEX

Page numbers in bold refer to tables. Page numbers in italics refer to figures. Page numbers with "n" refer to notes.

SUBJECT INDEX

Page numbers in bold refer to tables. Page numbers in italics refer to figures. Page numbers with "n" refer to notes.