Preface

The central aim of many studies in the physical, behavioral, social, and biological sciences is the elucidation of cause-effect relationships among variables or events. However, the appropriate methodology for extracting such relationships from data—or even from theories—has been fiercely debated.

The two fundamental questions of causality are: (1) What empirical evidence is required for legitimate inference of cause-effect relationships? (2) Given that we are willing to accept causal information about a phenomenon, what inferences can we draw from such information, and how? These questions have been without satisfactory answers in part because we have not had a clear semantics for causal claims and in part because we have not had effective mathematical tools for casting causal questions or deriving causal answers.

In the last decade, owing partly to advances in graphical models, causality has undergone a major transformation: from a concept shrouded in mystery into a mathematical object with well-defined semantics and well-founded logic. Paradoxes and controversies have been resolved, slippery concepts have been explicated, and practical problems relying on causal information that long were regarded as either metaphysical or unmanageable can now be solved using elementary mathematics. Put simply, causality has been mathematized.

This book provides a systematic account of this causal transformation, addressed primarily to readers in the fields of statistics, artificial intelligence, philosophy, cognitive science, and the health and social sciences. Following a description of the conceptual and mathematical advances in causal inference, the book emphasizes practical methods for elucidating potentially causal relationships from data, deriving causal relationships from combinations of knowledge and data, predicting the effects of actions and policies, evaluating explanations for observed events and scenarios, and—more generally—identifying and explicating the assumptions needed for substantiating causal claims.

Ten years ago, when I began writing *Probabilistic Reasoning in Intelligent Systems* (1988), I was working within the empiricist tradition. In this tradition, probabilistic relationships constitute the foundations of human knowledge, whereas causality simply provides useful ways of abbreviating and organizing intricate patterns of probabilistic relationships. Today, my view is quite different. I now take causal relationships to be the fundamental building blocks both of physical reality and of human understanding of that reality, and I regard probabilistic relationships as but the surface phenomena of the causal machinery that underlies and propels our understanding of the world.

Accordingly, I see no greater impediment to scientific progress than the prevailing practice of focusing all of our mathematical resources on probabilistic and statistical inferences while leaving causal considerations to the mercy of intuition and good judgment. Thus I have tried in this book to present mathematical tools that handle causal relationships side by side with probabilistic relationships. The prerequisites are startlingly simple, the results embarrassingly straightforward. No more than basic skills in probability theory and some familiarity with graphs are needed for the reader to begin solving causal problems that are too complex for the unaided intellect. Using simple extensions of probability calculus, the reader will be able to determine mathematically what effects an intevention might have, what measurements are appropriate for control of confounding, how to exploit measurements that lie on the causal pathways, how to trade one set of measurements for another, and how to estimate the probability that a one event was the actual cause of another.

Expert knowledge of logic and probability is nowhere assumed in this book, but some general knowledge in these areas is beneficial. Thus, Chapter 1 includes a summary of the elementary background in probability theory and graph notation needed for the understanding of this book, together with an outline of the developments of the last decade in graphical models and causal diagrams. This chapter describes the basic paradigms, defines the major problems, and points readers to the chapters that provide solutions to those problems.

Subsequent chapters include introductions that serve both to orient the reader and to facilitate skipping; they indicate safe detours around mathematically advanced topics, specific applications, and other explorations of interest primarily to the specialist.

The sequence of discussion follows more or less the chronological order by which our team at UCLA has tackled these topics, thus recreating for the reader some of our excitement that accompanied these developments. Following the introductory chapter (Chapter 1), we start with the hardest questions of how one can go about discovering cause-effect relationships in raw data (Chapter 2) and what guarantees one can give to ensure the validity of the relationships thus discovered. We then proceed to questions of identifiability—namely, predicting the direct and indirect effects of actions and policies from a combination of data and fragmentary knowledge of where causal relationships might operate (Chapters 3 and 4). The implications of these findings for the social and health sciences are then discussed in Chapters 5 and 6 (respectively), where we examine the concepts of structural equations and confounding. Chapter 7 offers a formal theory of counterfactuals and structural models, followed by a discussion and a unification of related approaches in philosophy, statistics, and economics. The applications of counterfactual analysis are then pursued in Chapters 8 and 9, where we develop methods of bounding causal relationships and illustrate applications to imperfect experiments, legal responsibility, and the probability of necessary, sufficient, and single-event causation. We end this book with a transcript of a public lecture (Chapter 11) that I presented at UCLA, which provides a gentle introduction of the historical and conceptual aspects of causation.

Readers who wish to be first introduced to the nonmathematical aspects of causation are advised to start with Chapter 11 and then to sweep through the other historical/conceptual parts of the book: Sections 1.1.1, 3.3.3, 4.5.3, 5.1, 5.4.1, 6.1, 7.1, 7.4, 7.5, 8.3, 9.1, 9.3, and 10.1.4. More formally driven readers, who may be anxious to delve directly into the mathematical aspects and computational tools, are advised to start with Section 7.1 and then to proceed as follows for tool building: Section 1.2, Chapter 3, Sections 4.2–4.4, Sections 5.2–5.3, Sections 6.2–6.3, Chapter 8, and Sections 9.2–10.

Acknowledgment

I owe a great debt to many people who assisted me with this work. First, I would like to thank the members of the Cognitive Systems Laboratory at UCLA, whose work and ideas formed the basis of many of these sections: Alex Balke, Blai Bonet, David Chickering, Adnan Darwiche, Rina Dechter, Hector Geffner, Dan Geiger, Moisés Goldszmidt, Jin Kim, Jin Tian, and Thomas Verma. Tom and Dan have proven some of the most basic theorems in causal graphs; Hector, Adnan, and Moisés were responsible for keeping me in line with the logicist approach to actions and change; and Alex and David have taught me that counterfactuals are simpler than the name may imply.

My academic and professional colleagues have been very generous with their time and ideas as I began ploughing the peaceful territories of statistics, economics, epidemiology, philosophy, and the social sciences. My mentors-listeners in statistics have been Phil Dawid, Steffen Lauritzen, Don Rubin, Art Dempster, and David Freedman. In economics, I have benefited from many discussions with John Aldrich, Kevin Hoover, James Heckman, Ed Leamer and Herbert Simon. My forays into epidemiology resulted in a most fortunate and productive collaboration with Sander Greenland and James Robins. Philosophical debates with James Woodward, Nancy Cartwright, Brian Skyrms, Clark Glymour, and Peter Spirtes have sharpened my thinking of causality in and outside philosophy. Finally, in artificial intelligence, I have benefited from discussions with and the encouragement of Nils Nilsson, Ray Reiter, Don Michie, and Joe Halpern.

The National Science Foundation deserves acknowledgment for consistently and faithfully sponsoring the research that led to these results, with special thanks to H. Moraff, Y. T. Chien and Larry Reeker. Other sponsors include Abraham Waksman of the Air Force Office of Scientific Research, Michael Shneier of the Office and Naval Research, the California MICRO Program, Northrop Corporation, Rockwell International, Hewlett-Packard, and Microsoft.

The manuscript was most diligently typed, processed, and illustrated by Kaoru Mulvihill. Jin Tian and Blai Bonet helped in proofing selected chapters. I thank the publisher for accommodating my idiosyncrasies, ... etc.

Finally, I owe a great debt to my family—to Tammy, Danny, Michelle, and Leora for filling my heart with their smiles, and to my wife Ruth for surrounding me with so much love, support, and meaning.

J.P.

Los Angeles

August 1999