Chapter 2

A THEORY OF INFERRED CAUSATION

I would rather discover one causal
law than be King of Persia.
Democritus (460–370 B.C.)

Preface

The possibility of learning causal relationships from raw data has been on philosophers' dream lists since the time of Hume (1711–1776). That possibility entered the realm of formal treatment and feasible computation in the mid-1980s, when the mathematical relationships between graphs and probabilistic dependencies came into light. The approach described herein is an outgrowth of Pearl (1998b, Chap. 8), which describes how causal relationships can be inferred from nontemporal statistical data if one makes certain assumptions about the underlying process of data generation (e.g., that it has a tree structure). The prospect of inferring causal relationships from weaker structural assumptions (e.g., general directed acyclic graphs) has motivated parallel research efforts at three universities: UCLA, Carnegie Mellon University (CMU), and Stanford. The UCLA and CMU teams pursued an approach based on searching the data for patterns of conditional independencies that reveal fragments of the underlying structure and then piecing those fragments together to form a coherent causal model (or a set of such models). On the other hand, the Stanford group pursued a Bayesian approach, where data are used to update the posterior probabilities assigned to candidate causal structures [Cooper and Herskovits, 1991]. The UCLA and CMU efforts have led to similar theories and almost identical discovery algorithms, which were implemented in the TETRAD II program [Spirtes et al., 1993]. The Bayesian approach has since been pursued by a number of research teams [Singh and Valtorta, 1995; Heckerman et al., 1994] and now serves as the basis for several graph-based learning methods [Jordan, 1998]. This chapter describes the approach pursued by Tom Verma and me in the period 1988–1992, and it briefly summarizes related extensions, refinements, and improvements that have been advanced by the CMU team and others. Some of the philosophical rationale behind this development, primarily the assumption of minimality, are implicit in the Bayesian approach as well (Section 2.9.1).

The basic idea of automating the discovery of causes—and the specific implementation of this idea in computer programs—came under fierce debate in a number of forums [Cartwright, 1995a; Humphreys and Freedman, 1996; Cartwright, 1997; Korb and Wallace, 1997; McKim and Turner, 1997; Robins and Wasserman, 1999]. Selected aspects of this debate will be addressed in the discussion section at the end of this chapter (Section 2.9.1).

Acknowledging that statistical associations do not *logically* imply causation, this chapter asks whether weaker relationships exist between the two. In particular, we ask:

1. What clues prompt people to perceive causal relationships in uncontrolled observations?

2. Is it feasible to infer causal models from these clues?

3. Would the models inferred tell us anything useful about the causal mechanisms that underly the observations?

In Section 2.2 we define the notions of causal models and causal structures and then describe the task of causal modeling as an inductive game that scientists play against Nature. In Section 2.3 we formalize the inductive game by introducing “minimal model” semantics—the semantical version of Occam’s razor—and exemplify how, contrary to common folklore, causal relationships can be distinguished from spurious covariations following this standard norm of inductive reasoning. Section 2.4 identifies a condition, called *stability* (or *faithfulness*), under which effective algorithms exist that uncover structures of casual influences as defined here. One such algorithm (called IC), introduced in Section 2.5, uncovers the set of all causal models compatible with the data, assuming all variables are observed. Another algorithm (IC*), described in Section 2.6, is shown to uncover many (though not all) valid causal relationships when some variables are *not* observable. In Section 2.7 we extract from the IC* algorithm the essential conditions under which causal influences are identified, and we offer these as independent definitions of genuine influences and spurious associations, with and without temporal information. Section 2.8 offers an explanation for the puzzling yet universal agreement between the temporal and statistical aspects of causation. Finally, Section 2.9 summarizes the claims made in this chapter, re-explicates the assumptions that leads to these claims, and offers new justifications of these assumption in light of ongoing debates.