2.1 Introduction

An autonomous intelligent system attempting to build a workable model of its environment cannot rely exclusively on preprogrammed causal knowledge; rather, it must be able to translate direct observations to cause-and-effect relationships. However, given that statistical analysis is driven by covariation, not causation, and assuming that the bulk of human knowledge derives from uncontrolled observations, we must still identify the clues that prompt people to perceive causal relationships in the data. We must also find a computational model that emulates this perception.

Temporal precedence is normally assumed to be essential for defining causation, and it is undoubtedly one of the most important clues that people use to distinguish causal from other types of associations. Accordingly, most theories of causation invoke an explicit requirement that a cause precedes its effect in time (Reichenbach 1956; Good 1961; Suppes 1970; Shoham 1988). Yet temporal information alone cannot distinguish genuine causation from spurious associations caused by unknown factors—the barometer falls before it rains yet does not cause the rain. In fact, the statistical and philosophical literature has adamantly warned analysts that, unless one knows in advance all causally relevant factors or unless one can carefully manipulate some variables, no genuine causal inferences are possible(Fisher 1951; Skyrms 1980; Cliff 1983; Eells and Sober 1983; Holland 1986; Gardenfors 1988; Cartwright 1989).[1] Neither condition is realizable in normal learning environments, and the question remains how causal knowledge is ever acquired from experience.

The clues that we explore in this chapter come from certain patterns of statistical associations that are characteristic of causal organizations—patterns that, in fact, can be given meaningful interpretation only in terms of causal directionality. Consider, for example, the following *intransitive* pattern of dependencies among three events: A and B are dependent, B and C are dependent, yet A and C are independent. If you ask a person to supply an example of three such

[1] Some of the popular quotes are: "No causation without manipulation" (Holland 1986), "No causes in, no causes out" (Cartwright 1989), "No computer program can take account of variables that are not in the analysis" (Cliff 1983).

events, the example would invariably portray A and C as two independent causes and B as their common effect, namely, $A \rightarrow B \leftarrow C$. (In my favorite example, A and C are the outcomes of two fair coins, and B represents a bell that rings whenever either coin comes up heads.) Fitting this dependence pattern with a scenario in which B is the cause and A and C are the effects is mathematically feasible but very unnatural (the reader is encouraged to try this exercise).

Such thought experiments tell us that certain patterns of dependency, which are totally void of temporal information, are conceptually characteristic of certain causal directionalities and not others. Reichenbach (1956) suggested that this directionality is a characteristic of Nature, reflective of the temporal asymmetries associated with the second law of thermodynamics. In Section 2.8 we offer a more subjective explanation, attributing the directionality to choice of language and to certain assumptions (e.g., Occam's razor and stability) prevalent in scientific induction. The focus of our investigation in this chapter is to explore whether this directionality provides a significant source of causal information and whether this information can be given formal characterization and an algorithmic implementation.

We start by introducing a model-theoretic semantics that gives a plausible account for how causal models could coherently be inferred from observations. Using this semantics we show that, subject to certain plausible assumptions, genuine causal influences can in many cases be distinguished from spurious covariations and, moreover, the direction of causal influences can often be determined without resorting to chronological information. (Although, when available, chronological information can significantly simplify the modeling task.)