by

Peter Spirtes[*¶] and Richard Scheines[*]

Among other things, causal hypotheses ought to predict how the world will respond to an intervention. How much will we reduce our risk of stroke by switching to a low-fat diet? How will the chances of another terrorist attack change if the U.S. invades Iraq next week? Causal inference is the move from data and background knowledge to justified causal hypotheses. Epistemologically, we want to characterize the conditions under which we can do causal inference, that is, what sorts of data and background knowledge can be converted into knowledge of how the world will respond to an intervention. Over the last two decades, philosophers, statisticians, and computer scientists have converged substantially on at least the fundamental outline of a theory of causation that provides a precise theory of causal knowledge and causal inference (Spirtes, Glymour, and Scheines, 2000; Pearl, 2000). Different researchers give slightly different accounts of the idea of a manipulation, or an intervention, but all assume that when we intervene ideally to directly set the value of exactly one variable, it matters not how we set it in predicting how the rest of the system will respond. This assumption turns out to be problematic, primarily because it often does matter how one sets the value of a variable one is manipulating. In this paper we explain the nature of the problem and what can be done to handle it. We begin

Consider the following hypothetical example. Through an observational study, researchers discover, they think, that high cholesterol levels cause heart disease. They recommend lower cholesterol diets to prevent heart disease. But, unknown to them, there are two sorts of cholesterol: LDL cholesterol causes heart disease, and HDL cholesterol prevents heart disease. Low cholesterol diets differ, however, in particular in the proportions of the two kinds of cholesterol. Consequently, experiments with low cholesterol regimens differ considerably in their outcomes.

In such a case the variable identified as causal—total cholesterol—is actually a deterministic function of two underlying factors, one of which is actually causal, the other preventative. The interventions (diets) are actually interventions on the underlying factors, but in different proportions. When specification of the value of a variable, such as total cholesterol, underdetermines the values of underlying causal variables, such as LDL cholesterol and HDL cholesterol, we will say that manipulation of that variable is ambiguous. How are such causal relations to be represented, what relationships between causal relations and probability distributions are there in such cases, and how should one conduct search when the systems under study may, for all one knows, have this sort of hidden structure? These issues seem important to

distribution over the values of , even those that have zero probability in the population distribution over , as long as the members of are jointly independent in the manipulated distribution. For a set of variables , a manipulation

In a causal

$P(HD = \text{Present}|HDL = \text{High}, LDL = \text{High}) =$
$P(HD = \text{Present}|TC = \text{High}) = .8$

In this case, while manipulating *TC* to Medium represents several different manipulations of the underlying variables *HDL* and *LDL*, each of the different manipulations of *HDL* and *LDL* compatible with manipulating *TC* to Medium produces the same effect on *HD* (i.e. $P(HD)$ after manipulation is equal to $P(HD = \text{Present}|HDL =$

4.3. Example 3

Examples 1 and 2 are two simple cases in which causal conclusions can be reliably made. Indeed, for those examples, the algorithms that we have already developed and that are reliable under the assumption that there are no ambiguous manipulations, still give correct output,