Causal Inference of Ambiguous Manipulations

by

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

1. Introduction

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 by describing the source of the problem, defined variables. We illustrate how interventions on defined variables can be ambiguous, and how this ambiguity affects prediction. We then describe how the possibility of ambiguous manipulations affects causal inference, and illustrate with an example involving both an ambiguous and then an unambiguous manipulation.

2. Defined Variables

In causal modeling, variables are sometimes deliberately introduced as defined functions of others variables. More interestingly, sometimes two or more measured variables are deterministic functions of one another, not deliberately, but because of redundant measurements, or underlying lawlike connections. This sort of dependency sometimes shows up as perfect correlation, also known as “multicollinearity,” which creates problems for data analysis for which a variety of strategies have been developed, e.g., “ridge regression.” Perhaps the most principled response is to divide the analysis into several sub-analyses in none of which are variables deterministically related. But the most interesting case is much more interesting.

[*] Department of Philosophy, Carnegie Mellon University. [¶] Institute for Human and Machine Cognition, University of West Florida.

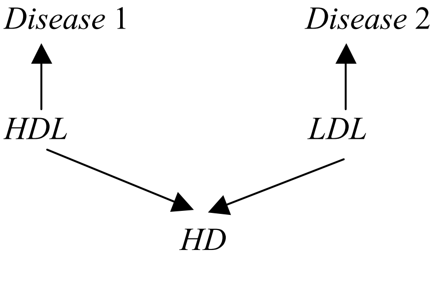
Consider the following hypothetical example. Through an observational study, researchers discover, they think, that high cholesterol levels cause heart disease. They recommend lower cholesterol diets to prevent heart disease. But, unknown to them, there are two sorts of cholesterol: LDL cholesterol causes 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 understanding possible reasons for disagreements between observational and experimental studies, non-repeatability of experimental studies (and not only in medicine—psychology present many examples), and in understanding the value and limitations of meta-analysis.

3. Causal Inference When Manipulations are Assumed Unambiguous

First, we will consider the case where all manipulations are assumed to be unambiguous. The general setup is described at length in Spirtes et al. (2000), which we illustrate with the following example. It is assumed that HDL cholesterol (*HDL*) causes *Disease* 1, LDL cholesterol (*LDL*) causes *Disease* 2, and that *HDL* and *LDL* cause heart disease (*HD*). This causal structure can be represented by the directed acyclic graph shown in

Figure 1.



HDL: High Density Lipids
LDL: Low Density Lipids
HD: Heart Disease

Figure 1

A directed graph G is the **causal graph** for a causal system C when there is an edge from A to B in G if and only if A is a direct cause of B relative to C. Note that for a causal system *C* there is a unique causal graph for *C*. (We assume that the set of variables in a causal graph is causally sufficient, i.e. if **V** is the set of variables in the causal graph, that there is no variable *L* not in **V** that is a direct cause (relative to $\mathbf{V} \cup \{L\}$) of two variables in **V**).

We assume that a causal system *C* over a set of causally sufficient variables **V** satisfies the Causal Markov Principle: each variable is independent of the set of variables that are neither its parents nor its descendants, conditional on its parents in the causal graph *G* for *C*. In this case, that entails that *HD* is independent of {*Disease* 1, *Disease* 2} conditional on {*HDL*, *LDL*}, *HDL* is independent of {*LDL*, *Disease* 2}, *LDL* is independent of {*HDL*, *Disease* 1}, *Disease* 1 is independent of {*HD*, *LDL*, *Disease* 2} conditional on *HDL*, and *Disease* 2 is independent of {*HD*, *HDL*, *Disease* 1} conditional on *LDL*.

The Causal Markov Principle is equivalent to the following factorization principle. *X* is an **ancestor** of *Y* in a graph if there is a directed path from *X* to *Y*, or *X* = *Y*. *X* is a **parent** of *Y* in a graph if there is a directed edge $X \rightarrow Y$ in the graph. **Parents**(*G*,*Y*) is the set of parents of *Y* in graph *G*. If *G* is a directed graph over **S**, and $\mathbf{X} \subseteq \mathbf{S}$, **X** is an **ancestral set** of vertices relative to *G* if and only if every ancestor of **X** in *G* is in **X**. A joint probability distribution (or in the case of continuous variables a joint density function) *P*(**V**) **factors according to a directed acyclic graph** (DAG) *G* when for every $\mathbf{X} \subseteq \mathbf{V}$ that is an ancestral set relative to *G*,

$$P(\mathbf{X}) = \prod_{X \in \mathbf{X}} P(X \mid \mathbf{Parents}(G,X))$$

In the example, this entails that *P*(*Disease* 1, *Disease* 2, *HDL*, *LDL*, *HD*} = *P*(*Disease* 1|*HDL*) × *P*(Disease 2|*LDL*) × *P*(*HDL*) × *P*(*LDL*) × *P*(*HD*|*LDL*,*HDL*). This factorization entails that the entire joint distribution can be specified in the following way:

P(*HDL* = High) = .2
P(*LDL* = High) = .4
P(*Disease* 1 = Present|*HDL* = Low) = .2
P(*Disease* 1 = Present|*HDL* = High) = .9
P(*Disease* 2 = Present|*LDL* = Low) = .3
P(*Disease* 2 = Present|*LDL* = High) = .8

P(*HD* = Present|*HDL* = Low, *LDL* = Low) = .4
P(*HD* = Present|*HDL* = High, *LDL* = Low) = .1
P(*HD* = Present|*HDL* = Low, *LDL* = High) = .8
P(*HD* = Present|*HDL* = High, *LDL* = High) = .3

We will take a **joint manipulation** $\{Man(P_1(X_1))\ldots,Man(P_n(X_n))\}$ for a set of variables $X_i \in \mathbf{X}$ as primitive. Intuitively, this represents a randomized experiment where the distribution $P'(\mathbf{X}) = \prod_{X_i \in \mathbf{X}} P'_i(X_i)$ is forced upon the variables **X**. We will also write $\{Man(P'_1(X_1))\ldots,Man(P'_n(X_n))\}$ as $Man(P'(\mathbf{X}))$. We assume **X** can be manipulated to any distribution over the values of **X**, even those that have zero probability in the population distribution over **X**, as long as the members of **X** are jointly independent in the manipulated distribution. For a set of variables $\mathbf{V} \supseteq \mathbf{X}$, a manipulation *Man*(*P'*(**X**)) transforms a distribution *P*(**V**) into a manipulated distribution over **V** denoted as *P*(**V**||*P'*(**X**)), where the double bar (Lauritzen??) denotes that the manipulation *Man*(*P'*(**X**)) has been performed.

A DAG *G* **represents** a causal system among a set of variables **V** when *P*(**V**) factors according to *G*, and for every manipulation *Man*(*P*(**X**)) of any subset **X** of **V**

$$P(\mathbf{V} \,||\, P'(\mathbf{X})) = P'(\mathbf{X}) \,\square\, \mathop{\square}\limits_{V \square \mathbf{V}\backslash\mathbf{X}} P(V \,|\, \mathbf{Parents}(G,V))$$

For example, if *HD* is manipulated to have the value Present, i.e. *P'*(*HD* = Present) = 1, then the joint manipulated distribution is:

P(*Disease* 1, *Disease* 2, *HD*, *LDL*, *HD* = Present||*P'*(*HD*)) = 1 □ *P*(*Disease* 1|*HDL*) □ *P*(*Disease* 2|*LDL*) □ *P*(*LDL*) □ *P*(*HDL*), and

P(*Disease* 1, *Disease* 2, *HDL*, *LDL*, *HD* = Absent||*P'*(*HD*)) = 0 □ *P*(*Disease* 1|*HDL*) □ *P*(*Disease* 2|*LDL*) □ *P*(*LDL*) □ *P*(*HDL*) = 0.

Note that the distribution of *HDL* after manipulation of *HD* to Present, i.e. *P*(*HDL*||*P'*(*HD* = Present) = 1) is not equal to the probability of *HDL* conditional on *HD* = Present, i.e. *P*(*HDL*|*HD* = Present).

The Causal Markov Principle allows some very limited causal inferences to be made. For example, suppose {*X*,*Y*} is causally sufficient. A causal graph that contains no edge between *X* and *Y* entails that *X* is independent of *Y*, and hence is not compatible with any distribution in which *X* and *Y* are dependent. However, if it is not known whether {*X*,*Y*} is causally sufficient, then assuming just the Causal Markov Principle, no causal conclusions can be reliably drawn. For example, suppose the causal relationships between *X* and *Y* is known to be linear, and the correlation between *X* and *Y* is *r*. Let the linear coefficient that describes the effect of *X* on *Y* be *c* (i.e. a unit change in *X* produces a change of *c* in *Y*.) Then regardless of the value of *r*, for any specified *c* there is a causal model in which the correlation between *X* and *Y* is *r*, and the effect of *X* on *Y* is *c*. That is the observed statistical relation (*r*) places no constraints at all on the causal relation (*c*). This negative result generalizes the case where more than two variables are measured. Even when {*X*,*Y*} is known to be causally sufficient, the Causal Markov Principle does not suffice to produce a unique prediction about the mean effect of manipulating *X* on *Y* no matter what variables are measured, as long as *X* and *Y* are dependent.

We will make a second assumption, that is commonly, if implicitly, made in the statistical literature.

Causal Faithfulness Principle: In a causal system C, if **S** is causally sufficient, and *P*(**S**) is the distribution over **S** in C, every conditional independence that holds in *P*(**S**) among three disjoint sets of variables **X**, **Y**, and **Z** included in **S** is entailed by the causal graph that represents C under the Causal Markov Condition.

The justification for the Causal Faithfulness Principle (as well as descriptions of cases where it should not be assumed) is discussed at length in Spirtes et al. (2000). One justification is that for a variety of parametric families, the Causal Faithfulness Principle is only violated for a set of parameters that have measure 0 (with respect to Lebesgue measure, and hence with respect to any of the usual priors placed over the parameters of the model.)

Given the Causal Markov Principle, and the Causal Markov Principle, there are algorithms that in the large sample limit correctly infer some of the causal relations among the random variables, and correctly predict the effects of some manipulations, even if it is not known whether the measured variables are causally sufficient. For those causal relations that cannot be inferred, and those effects of manipulations that cannot be predicted, the algorithms will return “can’t tell”.

4. Causal Inference When Manipulations May Be Ambiguous

Consider what kinds of dependency structures can emerge in a few hypothetical examples.

4.1. Example 1

Consider an extenstion of the hypothetical Example 1, shown in Figure 2, in which the concentration of total cholesterol is defined in terms of the concentrations of high density lipids and low density lipids. This is indicated in the figure by the bold faced arrows from *HDL* and *LDL* to *TC*. The other arrows indicate causal relationships. Suppose that high levels of *HDL* tend to prevent *HD*, while high levels of *LDL* tend to cause *HD*. We have the following parameters for Example 1.

HDL = Low, *LDL* = Low □ *TC* = Low

HDL = Low, *LDL* = High □ *TC* = Medium

HDL = High, *LDL* = Low □ *TC* = Medium

HDL = High, *LDL* = High □ *TC* = High

P(*HDL* = High) = .2

P(*LDL* = High) = .4

P(*Disease* 1 = Present|*HDL* = Low) = .2

P(*Disease* 1 = Present|*HDL* = High) = .9

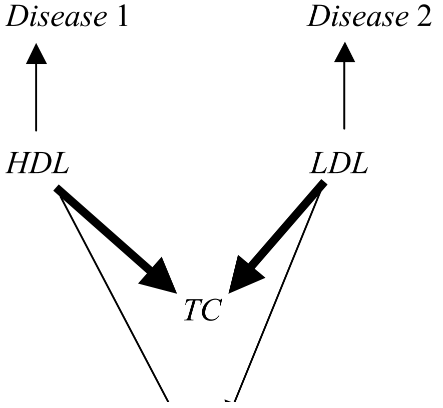
P(*Disease* 2 = Present|*LDL* = Low) = .3

P(*Disease* 2 = Present|*LDL* = High) = .8

$P(HD = \text{Present}|HDL = \text{Low}, LDL = \text{Low}) = .4 = P(HD = \text{Present}|TC = \text{Low})$
$P(HD = \text{Present}|HDL = \text{High}, LDL = \text{Low}) = .1$
$P(HD = \text{Present}|HDL = \text{Low}, LDL = \text{High}) = .8$
$P(HD = \text{Present}|HDL = \text{High}, LDL = \text{High}) = .3 = P(HD = \text{Present}|TC = \text{High})$

Manipulation of *TC* is really a manipulation of *HDL* and *LDL*. However, even after an exact level of *TC* is specified as the target of a manipulation, there are different possible manipulations of *HDL* and *LDL* compatible with that target. For example, if a manipulation sets *TC* to Medium, then this could be produced by manipulating *HDL* to Low and *LDL* to High, or by manipulating *HDL* to High and *LDL* to Low. Thus, even after the manipulation of *TC* is completely specified (e.g. to Medium), the *effect* of the manipulation on *HD* is indeterminate (i.e. if the manipulation is *HDL* to High and *LDL* to Low, then after the manipulation *P*(*HD*) is .1, but if the manipulation is *HDL* to Low and *HDL* to High, then after the manipulation *P*(*HD*) is .8). Hence a manipulation of *TC* to Medium might either lower the probability of *HD* (compared to the population rate), or it might raise the probability of *HD*. It is quite plausible that in many instances, someone performing a manipulation upon *TC* would not know about the existence of the underlying variables *HDL* and *LDL*, and would not know that the manipulation they performed was ambiguous with respect to underlying variables. For example, manipulation of *TC* could be produced by the administration of several different drugs that affect *HDL* and *LDL* in different ways, and produce different effects on *HD*.

What is the correct answer to the question "What is the effect of manipulating *TC* to Medium on *HD*?" With no further information, the most informative answer that could be given is to give the entire range of effects of manipulating *TC* to Medium (i.e. either *P*(*HD*) = .8 or *P*(*HD* = .1)). One might take a Bayesian strategy in which priors were put over the probability of the underlying variables from which the manipulated variable is constructed. Another possible answer is to simply output "Can't tell" because the answer is indeterminate from the information given. A fourth, but misleading, answer would be to output one of the many possible answers (e.g. *P*(*HD*) = .1).This answer is misleading as long as it contains no indication that this is merely one of a set of possible different answers, and an actual manipulation of *TC* to Medium might lead to a completely different result. Note that it is the third, misleading, kind of answer that would be produced by performing a randomized clinical trial on *TC*; there would be nothing in the trial to indicate that the results of the trial depended crucially upon details of how the manipulation was done.



HDL: High Density Lipids
LDL: Low Density Lipids
TC: Total Cholestorol
HD: Heart Disease

Figure 2

Suppose now that *Disease* 1, *Disease* 2, *TC*, and *HD* are the measured variables, and we assume the Causal Markov and Faithfulness Principles (extended to graphs with definitional links), but allow that there may be hidden common causes. What reliable (pointwise consistent) inferences can be drawn from samples of the distribution described in Example 1? We will contrast 2 cases: the case where it is assumed that all manipulations are unambiguous and the case where the possibility that a manipulation may be ambiguous is allowed. The general effect of weakening the assumption of no ambiguous manipulations is to introduce more "Can't tell" entries.

Manipulate:	Effect on:	Assume manipulation unambiguous	Manipulation may be ambiguous
Disease 1	*Disease* 2	None	None
Disease 1	*HD*	Can't tell	Can't tell
Disease 1	*TC*	Can't tell	Can't tell
Disease 2	*Disease* 1	None	None
Disease 2	*HD*	Can't tell	Can't tell
Disease 2	*TC*	Can't tell	Can't tell
TC	*Disease* 1	None	Can't tell
TC	*Disease* 2	None	Can't tell
TC	*HD*	Can't tell	Can't tell
HD	*Disease* 1	None	Can't tell
HD	*Disease* 2	None	Can't tell
HD	*TC*	Can't tell	Can't tell

4.2. Example 2

We will now consider what happens when the example is changed slightly. In Example 2, suppose that the effect of *HDL* and *LDL* on the probability of *HD* actually is completely determined by *TC*. Example 2 is the same as the Example 1, except that we have changed the distribution of *HD* in the following way:

$P(HD = \text{Present}|HDL = \text{Low}, LDL = \text{Low}) =$
$P(HD = \text{Present}|TC = \text{Low}) = .1$

$P(HD = \text{Present}|HDL = \text{High}, LDL = \text{Low}) =$
$P(HD = \text{Present}|HDL = \text{Low}, LDL = \text{High}) =$
$P(HD = \text{Present}|TC = \text{Medium}) = .3$

P(*HD* = Present|*HDL* = High, *LDL* = High) =
P(*HD* = Present|*TC* = 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* = Present|*HDL* = High, *LDL* = Low) = *P*(*HD* = Present|*HDL* = Low, *LDL* = High) prior to manipulation, which is .3). In this case we say that the effect of manipulating *TC* on *HDL* is *determinate*. (Note that the effect of manipulating *TC* on *Disease* 1 is not determinate, because it depends upon how the manipulation of *TC* is done. So manipulating a variable may have determinate effects on some variables, but not on others.)

Interestingly, the Causal Faithfulness assumption actually entails that the effect of *TC* on *HD* is not determinate. This is because if the effect of manipulating *TC* on *HD* is determinate, then *LDL* and *HDL* are independent of *HD* conditional on *TC*, which is not entailed by the structure of the causal graph, but instead holds only for certain values of the parameters, i.e. those values for which P(*HD* = Present|*HD* = Low, *LDL* = High) = P(*HD* = Present|*HDL* = High, *LDL* = Low). Hence, in these cases we make a modified version of the Causal Faithfulness Principle, which allows for the possibility of just these kinds of determinate manipulations.

What reliable (pointwise consistent) inferences can be drawn from samples of the distribution described in Example 2? Because there are conditional independence relations that hold in Example 2 that do not hold in Example 1, more pointwise consistent estimates of manipulated quantities can be made under the assumption that manipulations may be ambiguous, than could be made in the previous example.

Manipulate:	Effect on:	Assume manipulation unambiguous: Example 2	Manipulation may be ambiguous: Example 2
Disease 1	*Disease* 2	None	None
Disease 1	*HD*	Can’t tell	Can’t tell
Disease 1	*TC*	Can’t tell	Can’t tell
Disease 2	*Disease* 1	None	None
Disease 2	*HD*	Can’t tell	Can’t tell
Disease 2	*TC*	Can’t tell	Can’t tell
TC	*Disease* 1	None	Can’t tell
TC	*Disease* 2	None	Can’t tell
TC	*HD*	= P(*HD*|*TC*)	= P(*HD*|*TC*)
HD	*Disease* 1	None	None
HD	*Disease* 2	None	None
HD	*TC*	None	None

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, as long as the output is suitably reinterpreted according to some simple rules that only slightly weaken the conclusions that can be drawn. However, there are other examples in which this is not the case. For example, if *Disease* 1 and *Disease* 2 are not independent, but are independent conditional on a third measured variable *X* then no simple reinterpretation of the output of the algorithm gives answers which are both informative about cases in which *TC* does determinately cause *HD*, and reliable. In all such examples that we have examined so far, however, the data itself contains information which indicates that the current algorithm cannot be applied reliably; hence for these examples the algorithm could simply be modified to check the data for this condition, and output "can't tell."

We do not yet have general conditions under which the data would indicate that the algorithm cannot be reliably applied (unless the assumption of no ambiguous manipulations is made.) This raises the question: Are there feasible general algorithms that are both correct and informative even when the assumption of no ambiguous manipulations is not made? If so, what are they? What kind of computational complexity as a function of the number of variables will such algorithms require? What sorts of sample sizes will such algorithms require in order to be useful?

5. References

Lauritzen, S. ().

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press.

Spirtes, P., Glymour, C., and Scheines, R. (2000) *Causation*, *Prediction*, *and Search*. MIT Press, Cambridge MA.