PHILOSOPHICAL TOPICS
VOL. 18 NO. 2, FALL 1990

*Independence Relations Produced by Parameter Values in Causal Models**

Clark Glymour, Peter Spirtes, and Richard Scheines
Carnegie Mellon University

I

INTRODUCTION

In *Discovering Causal Structure* (Glymour et al. 1987), we began the project of developing a theoretical and practical basis for computerized inference from empirical patterns of constraints on covariances to directed graphs representing causal structure. That work built on a history of prior treatments of particular classes of linear causal models in the engineering, biometric, psychometric, and social scientific literature. Judea Pearl and his associates (Pearl 1988; Verma and Pearl 1990), and before them several others (Kiiveri and Speed 1982; Wermuth and Lauritzen 1983), produced a general account, not confined to linear models, of how directed acyclic graphs representing causal structure also represent relations of statistical independence. In a series of recent papers, we have applied these representations to obtain an algorithm that automatically builds a family of directed acyclic graphs from the independence relations of any joint probability distribution on a set of measured variables, assuming there are no unmeasured common causes (Spirtes 1990, Spirtes forthcoming). We have implemented the procedure in the TETRAD II program for discrete and continuous variables. Verma and Pearl have shown that for any probability distribution having a perfect representation by some directed acyclic graph the procedure generates all and only the graphs that represent the distribution (Verma and Pearl 1990). This enables a characterization of the equivalence class of any graph, i.e., of the graphs that perfectly represent the same independence relations. These and related results seem to open the prospect of reliable, feasible procedures for extracting causal conclusions from statistical data.

Nancy Cartwright has recently published a criticism of the project (Cartwright 1989), founded on the observation that, in linear models, relations of statistical independence can arise because of special values of linear parameters. We wish to consider objections of this kind.

Relations of conditional independence can arise among variables in a causal structure either because of the causal structure itself, or because special forms of conditional probability happen to obtain. In some forms of causal theory, the conditional probabilities are determined by the directed graph of causal relations and by an underlying set of parameters; special values of the parameters will interact with the causal structure to produce conditional independencies that do not result from the causal structure alone. We will illustrate this with examples that are *linear* causal models.
A linear causal model consists of:

(i) a set S of random variables,
(ii) a set of causal relations between variables in S,
(iii) a set of distributions over the exogenous variables (variables that have no cause in S), and
(iv) a set of linear equations relating each endogenous variable (variables that have a cause in S) to other variables in S.

The distributions of the exogenous variables and the linear equations together generate a joint probability distribution over the variables in S.

The causal structure of a linear causal model is represented by a directed graph in which there is an arrow from A to B if and only if A is a direct cause of B. Figure 1 is an example of a graph that represents the causal structure of a linear causal model.

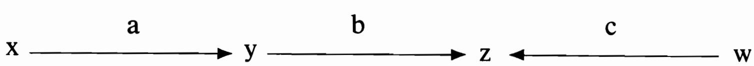

FIGURE 1

In Figure 1, X is a direct cause of Y, but not of Z. We assume that all exogenous variables (variables of indegree 0 in the graph) are independent of each other[1], and that all endogenous variables (variables of indegree greater than 0 in the graph) are linear functions of their immediate causal ancestors. Thus the graph describes the *form* of the equations relating the variables, but it does not describe the actual distributions of the independent variables, nor the values of the linear coefficients. If, in addition to the graph itself, there is a label on an edge from A to B in the graph, it represents the coefficient of B in the linear equation for A. In Figure 1, $Z = bY + cW$.[2] In Figure 1, X and Z are independent conditional on Y regardless of the distribution of the exogenous variables, and regardless of the values of the linear coefficients *a, b,* and *c*.[3] We say that such a conditional independence relation is implied by the causal structure, since in *all* of the probability distributions generated by linear causal models with the causal structure depicted in Figure 1, X and Z are independent conditional on Y.

On the other hand, linear models can also imply conditional independence relations for particular values of their linear coefficients. In the linear model depicted in Figure 2(1), $x1$ and $x3$ are statistically independent only when the values of the linear coefficients *a, b,* and *c* satisfy $a = -bc$. In this case we say that the independence of $x1$ and $x3$ is due to special parameter values.

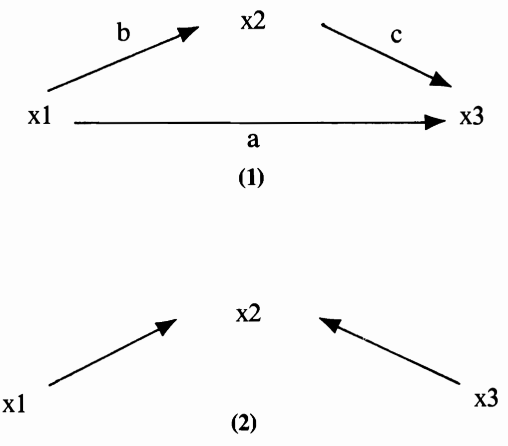

FIGURE 2

The model depicted in Figure 2(2) implies by virtue of its causal structure alone all of the conditional independencies implied by Figure 2(1); unlike the model depicted in Figure 2(1), Figure 2(2) does not need any special parameter values to imply these conditional independencies. We infer causal structure by searching for models that imply by virtue of their causal structure alone all conditional independence constraints that we judge to hold in the population. Hence, if the true causal model was the one depicted in Figure 2(1) with $a = -bc$, we would be fooled into believing that it was the model depicted in Figure 2(2).

One of the fundamental methodological ideas of *Discovering Causal Structure*, which we called "Spearman's principle" and attributed to the psychologist Charles Spearman, is that in causal inference one should not worry about the possibility that the causal structure will appear to be other than it is because of special values of the parameters. In defense of the principle, we cited a history of practice in physics and astronomy. These are, however, no more than arguments from (good) authority and a history of practice. They are not demonstrations that inference procedures that use Spearman's principle are reliable.

In her recent book, Cartwright objects that, since in linear models independence relations may be produced by special values of the linear coefficients and variances as well as by the causal structure, it is illegitimate to infer causal structure from such dependencies:

> Theorems like [those of] Glymour, Scheines, Kelly, and Spirtes . . . may seem to provide a way to mark out certain relations as relevant: the vanishing three-variable partial correlation is relevant to judging every [causal] structure because it implies some facts about what any [causal] structure will look like that can account for it. But the original question is always looming: should a [causal] structure be required to account for it? The model should do so, of course; but should it do so on the basis of [causal] structure alone? (p. 85)

One answer to the general criticism Cartwright offers is to defend Spearman's principle by showing that, under any of a family of natural measures on the space of parameters, the measure of those parameter values that will produce statistical independence relations not due to causal structure is zero. We have obtained theorems of this sort (Spirtes forthcoming). They show that independence relations produced by special parameter values are unstable. Any sufficiently small continuous change in the values of the parameters will eliminate the independencies their special values produce. As Brian Skyrms has pointed out, in social or other systems in which the parameters are not perfectly constant over time, longitudinal studies could eliminate the possibilities that concern Cartwright.

We propose, however, to leave such replies to one side and instead to consider directly the question of parameter induced independencies:

what indeterminancies in causal inference may arise among linear models because of statistical independence relations produced by special parameter values? We do not claim to solve the question completely, and we leave open a number of very interesting issues, including the problem of giving necessary and sufficient conditions. For reasons that will appear subsequently, we conjecture that for all linear causal stuctures, if the time order of the variables is known, then the occurrence of independence relations produced by special parameter values is always "given away" by the data. Finally, we will extract some irony from the fact, which we will prove, that Cartwright's alternative construction of the relation between probability and causality tacitly requires Spearman's principle. We begin by considering some general results that are not restricted to linear models.

II

CHARACTERIZATIONS OF THE PERFECT REPRESENTATION OF STATISTICAL INDEPENDENCE RELATIONS BY DIRECTED ACYCLIC GRAPHS

Consider a probability distribution P on a set U of variables, and consider a directed acyclic graph G having U as its vertex set. We may think of G as a representation of the causal relations among the elements of U, in which a directed edge $u1 \rightarrow u2$ indicates that $u1$ has an effect on $u2$ that is not mediated by any other variable in U. Assume that for every pair $u1$, $u2$ of variables in U, every common cause of $u1$, $u2$ is also in U. P determines a collection of facts about which pairs of variables are independent conditional on which subsets of U. What should be the relation between properties G and facts about conditional independence?

We say that a vertex u in a directed acyclic graph G is a *collider* on undirected path p in G if and only if p contains two edges directed into u. Following Pearl, we say that variables x, y are *d-separated* by set S if and only if there exists no undirected path p between x and y, such that (i) every collider on p has a descendent in S and (ii) no other vertex on p is in S. We say that x, y, are *d-connected* with respect to S if and only if they are not d-separated with respect to S. We say that two sets X, Y of variables are d-separated by S if and only if every pair $<x,y>$ in the cartesian product of X and Y is d-separated by S.

Let us say that $I(X,Z,Y)_P$ when X is independent of Y given Z in distribution P. Let us say that $D(X,Z,Y)_G$ when X is d-separated from Y by Z in graph G.

> Pearl's Representation Definition: A graph G *perfectly represents* a probability distribution P just when for all sets X, Y, Z of vertices, $D(X,Z,Y)_G$ iff $I(X,Z,Y)_P$.

An investigation of the properties of this notion of representation may be found in Pearl's book.

An illustration of d-connectedness is given in Figure 3.

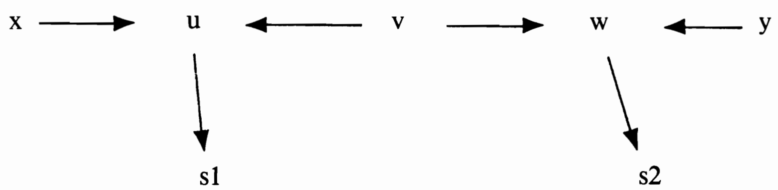

FIGURE 3

X and Y are not d-connected with respect to the empty set
X and Y are d-connected with respect to the set {S1, S2}
X and Y are not d-connected with respect to the set {S1, S2, V}

X and Y are independent
X and Y are not independent conditional on {S1, S2}
X and Y are independent conditional on {S1, S2, V}

Think of a mechanical device or electrical circuit arranged so that the variables in the graph have the causal relations illustrated. If you know the value of X and nothing else, it provides you with no information about the value of Y. If you know values of $S1$ and $S2$, then information about the value of X will give you additional information about the value of Y. If you know the value of $S1$, $S2$, and V—or even just the value of V, then information as to the value of X tells you nothing further about the value of Y.

A probability distribution may be perfectly represented by several distinct graphs. For example, in a distribution over U for which every pair of variables is dependent conditional on every subset of U not containing that pair, any acyclic orientation of the complete graph on U represents the distribution.

What is the connection between causal structure of a set of variables S and d-separability? If A and B are d-separated by C in the graph of the causal structure of a causal model M, then A and B are independent conditional on C in the probability distribution P generated by M (Pearl 1988). Of course, the causal graph is not necessarily a perfect representation of P, since there may be additional conditional independence relations true of P because of the particular parameter values of M. If there are no additional conditional independence relations true of P because of the particular parameter values of M, then the causal graph will be a perfect representation of P.

Suppose for the moment that we put aside the problem of independence relations holding due to special parameter values, and suppose that we have measured the values of all variables that are common causes of variables in S.[4]

Our Graph Construction Algorithm constructs the set of all graphs that perfectly represent a given set of independence relations over the set S of random variables (if the set is representable by a directed graph):

1. Add an undirected edge from A to B iff A and B are dependent given every subset of S not containing A and B.
2. If there are undirected edges between A and B, and B and C, but not between A and C, then there is an edge from A to B and from C to B iff A and C are dependent given every subset of S containing B, but not A or C.
3. The set of edges whose orientations are not fixed by 2 are given every possible orientation that does not create a collision ($A \rightarrow B \leftarrow C$).

The correctness of this algorithm follows from a theorem stated in Verma (1990). The complexity of the procedure is a polynomial function of the number of independence facts, but in the worst case the number of such facts is an exponential function of the number of variables. Such a procedure has been implemented in the TETRAD II program. The program takes in correlations or covariances, and tests hypotheses about vanishing partial correlations (of all orders). The vanishing partial correlations are then used as independence facts. The inference from vanishing partial correlation to conditional independence is of course perfectly rigorous if the joint distribution is normal. When the time order of the variables is known, the directed acyclic graph produced by this procedure is (ignoring sampling error) unique. The TETRAD II procedures represent a generalization of the procedures Cartwright questions, and presumably she would apply the very same criticism to them.

Pearl's connection between directed graphs and independence relations applies exactly to linear causal models. If the model specifies that two "error terms" are correlated, introduce a new variable that is a common cause of the variables associated with the error terms. The conversion procedure is discussed in detail in *Discovering Causal*

Structure. If one then applies Pearl's definition to the directed graph of the causal structure, one obtains only independence relations that hold in the distribution of the model. It may not, however, be the case that the d-separability criterion yields *all* of the independence relations that hold in the distribution of the model; other independence relations may hold because the linear coefficients and variances have special numerical values. Whence our subject.

III

PARAMETERS AND INDEPENDENCE

In linear models the variances of the exogenous variables (including error terms) and the linear coefficients associated with each directed edge are natural parameters. They, together with the graph structure, completely determine the covariances and hence in normal families the independence relations.

Parameter values cannot make a conditional independence implied by the graph of a model disappear. Parameter values can, however, force conditional independencies that are not implied by the true causal graph, as the example of the first section illustrates. There are two cases: if we start with a directed acyclic graph (DAG) and force extra independencies, the result may not be a probability distribution that can be represented by any DAG, or it may be a probability representation that is perfectly represented by some DAG, but not by a graph that correctly represents the causal structure. We first consider the following question:

> If graph G perfectly represents probability distribution P with conditional independence relations $C(P)$, and P' is a distribution perfectly represented by graph H and $C(P')$ contains $C(P)$, what is the relation between G and H?

The following propositions are easy consequences of the preceding results:

> **Proposition 1:** If graph G perfectly represents probability distribution P with conditional independence relations $C(P)$, and P' is a distribution on the same variables perfectly represented by graph H, then $C(P')$ contains $C(P)$ only if for all vertices x, y, if x, y are adjacent in H, they are adjacent in G.

> **Proposition 2:** If graph G perfectly represents probability distribution P with conditional independence relations $C(P)$, and P'

is a distribution on the same vertices perfectly represented by graph H, then $C(P')$ contains $C(P)$ only if (i) for every triple x, y, z of vertices such that $x \rightarrow y \leftarrow z$ is in G and x is not adjacent to z in G, if x, y are adjacent in H and y, z are adjacent in H, then $x \rightarrow y \leftarrow z$ in H, and (ii) for every triple x, y, z of vertices such that $x \rightarrow y \leftarrow z$ is in H and x is not adjacent to z in H, either $x \rightarrow y \leftarrow z$ in G or x, z are adjacent in G.

Proposition 3: If graph G perfectly represents probability distribution P with conditional independence relations $C(P)$, and P' is a distribution on the same variables perfectly represented by graph H, and G and H agree on a partial ordering of the variables (as by time) such that $x \rightarrow y$ only if $x < y$ in the partial order, and $C(P)$ contains $C(P')$, then H is a subgraph of G.

These results provide a bound on what parameters can do in the general case to mislead us about causal structure. Propositions 1 and 2, for example, tell us that the example depicted in Figure 2 represents the only way in which a mistake in orientation can occur so that two edges that in fact do not collide are taken to do so. If in that example the directed edges from $x1$ to $x2$ and from $x2$ to $x3$ are replaced by directed paths, then no illusion of a collision can be produced by special parameter values. If there is some variable $x4$ with an edge directed into $x2$ but not adjacent to $x1$, then no illusion of a collision can be produced by special parameter values.

IV

LINEARITY AND TIME ORDER

We are interested in two sorts of restrictions or "background knowledge," namely a time ordering of the variables and the assumption of linearity.

Proposition 3 says that if an ordering—as by time—of the variables is specified and variable x can cause variable y only if $x < y$, then if H is any graph consistent with the time order that perfectly represents a superset of the conditional independence relations represented by G, H is a sub-graph of G. This means that, when the time order of the variables is known, conditional independence relations due to special parameter values can only produce erroneous causal inferences in which a true connection is *omitted;* no other sorts of error may arise. For example, the sort of error illustrated in the first section of this paper cannot occur.

In linear models we assign a non-zero real number to each directed edge of a graph. These parameter values and the graph structure determine the correlations among the variables. In linear models it is routinely assumed that there is a set E (the "error variables") of variables of outdegree one and indegree zero such that every dependent variable not in E is adjacent to a variable in E. Generally, all variables including the error variables are required to have non-zero variance, although we will be forced to relax this requirement in what follows.

A *trek* is a pair of directed paths having a single common vertex that is the source of both paths (one of the paths in a pair may be the empty path). For standardized models in which the mean of each variable is zero and non-error variables have unit variance, the correlation of two variables is given by the sum over all treks connecting x, y of the product for each trek of the linear coefficients associated with the edges in that trek (we call this quantity the *trek sum*). We will use standardized models throughout our examples.

The system of correlations determines all partial correlations of every order through standard recursion relations. In a graph perfectly representing a probability distribution for a linear model, a conditional independence obtains only if the corresponding partial correlation vanishes (if the distribution is multinormal the converse obtains as well). Since the recursion relations give the same partial correlation no matter what the order of the variables conditioned on, a vanishing partial correlation corresponds to a system of equations in the coefficients of a standardized model.

Suppose now that special values of the linear parameters in a normal, standardized model G produce a system of partial correlations that is perfectly represented only by false causal structure, say H. Then the parameter values must require extra independence relations not represented by G; hence some partial correlation must vanish that does not correspond to a conditional independence represented by G. Any partial correlation is a function just of the trek sums connecting pairs of variables, and the trek sums in this case involve just the linear parameters in G. Hence each additional vanishing partial correlation not perfectly represented by G determines a system of (non-linear) equations in the parameters of G that the parameter values must satisfy in order to produce the "spurious" partial correlation. Now for some G and some H, a sub-graph of G, these systems of equations may have no simultaneous solution. In that case, there are no values for the parameters of G that will produce partial correlations that can be perfectly represented by H. For other choices of G and a subgraph H, it may be that the system of equations has a solution, but only solutions that allow only a finite number of alternative values for one or more parameters. For example, the alternative solution may imply that some correlation is equal to 1 or -1. If these parameters determine a correlation coefficient, then such a solution must "give itself away" by special correlation constraints that are not themselves conditional independence relations but rather follow from conditional independence relations only because of the true causal structure, G. Consider the following choices of G and H, where in each pair G is on the left hand side and H is on the right hand side (see Fig. 4).

In the first and last pairs, coefficients can be chosen for the graph on the left hand side so that it appears as though the edge does not occur, but only by making the coefficient labeled b equal to either 1 or -1. Since the variables are standardized, this requires that the error term for variable 2 have zero variance and zero mean—i.e., it vanishes. Thus in order for the true graph to be that on the lefthand side and the parameter values to produce independence relations that are perfectly represented by the graph on the righthand side, variable 2 must literally have the same value as does variable 1 for each unit in the (standardized) population. The same result obtains if the edges that are not eliminated in the first and last examples are replaced by directed paths of any length. Clearly in these cases special parameter values that produce independence relations not perfectly represented by the true graph must give themselves away. In the two middle cases, the edge between variables 1 and 3 cannot be made to appear to be eliminated by any choice of parameter values for the true graph. In fact the following principles hold:

> If G contains a directed edge from variable 1 to variable 2, and G contains two other treks between 1 and 2 that intersect only at 1 and 2, then no values of the parameters in G give independence relations perfectly represented by a graph H without a directed edge from 1 to 2 (Fig. 4.3).
>
> If G contains a directed edge from variable 1 to variable 2, and G contains a trek t between 1 and 2 and a variable j such that there is a directed edge from j to a vertex of t other than 1 or 2 and there is no other path from j to t, then no values of the parameters in G give independence relations perfectly represented by a graph H without a directed edge from 1 to 2 (Fig. 4.2).

In every pair G, H we have considered such that H is a subgraph of G and special values of the parameters of G determine partial correlations perfectly represented by H, there are extra constraints on the correlations not entailed by the partial correlations perfectly represented by H. We conjecture that this is always the case. In other

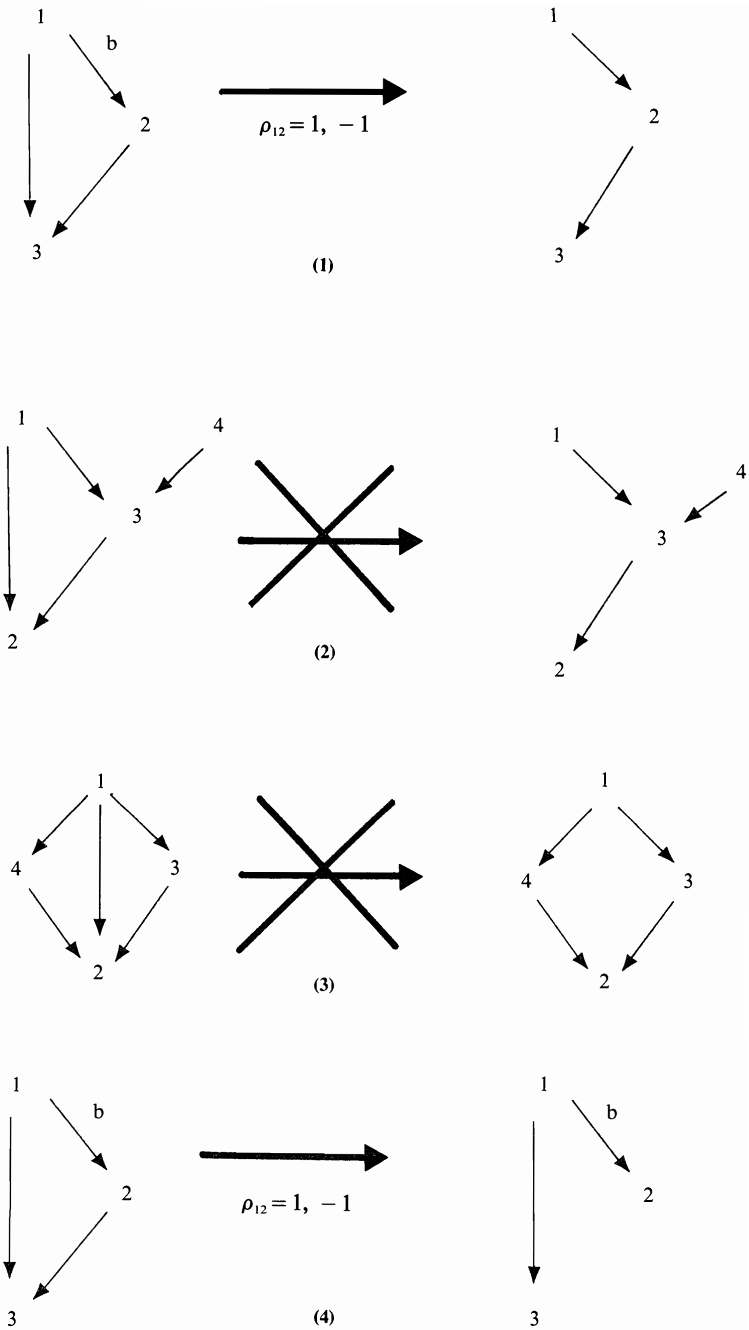

FIGURE 4

words, if time order is known, special parameter values cannot deceive us about the causal structure without giving themselves away. Further, unless three edges form a triangle as in the example in the first section, we conjecture that if parameter values of *G* determine a collection of partial correlations perfectly represented by a graph *H*—whether or not *H* is a subgraph of *G*—then there are extra constraints on the correlations not entailed by the partial correlations perfectly represented by *H*.

V

UNREPRESENTABLE CONSTRAINTS

We have considered the circumstances in which parameter values for a graph *G* can produce partial correlations perfectly represented by a graph *H*. What of the case in which values of the parameters of *G* produce a system *S* of partial correlations corresponding to conditional independence relations that cannot be perfectly represented by any directed acyclic graph? In that case, one could consider the maximal subsets of *S* that do have perfect representations. Any such subset may be perfectly represented by a graph *H* not identical with *G*, or by *G*. In the former case, the partial correlations of any maximal representable subset of *S* must have been produced from *G* by special parameter values, hence the preceding sections apply. Save where there is a triangle in *G* two of whose edges form collisions in *H* but not in *G*, if the conjectures of the previous section are correct, the correlation values will mark this indeterminacy.

In practice, the strategy of considering the maximal representable subsets of the set of independence relations found empirically is likely to be computationally infeasible even for small sets of variables. Heuristic methods that are asymptotically correct, give good results over typical cases on realistic sample sizes, and are computationally tractable, will be preferred in practice. Methods of this kind have been implemented in the TETRAD II program for elaborating models with latent variables. We plan to implement them as well for the task of constructing causal relations directly from the covariances.

VI

APPENDIX: SPEARMAN'S REVENGE

Spearman's principle is so natural that one often uses it without notice. To illustrate the point, we will show that Cartwright requires the principle for the central technical argument in the second chapter of her book. She makes the following claim, and offers a proof, which we will not analyze:

> Suppose there is a set of factors x_i, i = 1, . . . , m, each of which may, for all that is known, be a true cause of x_e, and such that
>
> $$S: x_e = \Sigma a_i x_i$$
>
> where each x_i has an open back path. Assume transitivity of causality. . . . Assume in addition . . . given any linear equation in the form $z_e = \Sigma b_i z_i$ which represents a true functional dependence, either (a) every factor on the righthand side is a true cause of the factor on the lefthand side, or (b) this equation can be derived from a set of equations for each of which (a) is true. Then each x_i is a true cause of x_e.

We take it that she also means to assume that in *S* none of the x_i are variables in that cause x_e only through some other x_j occurring in the sum.

Because of the phrase "for all that is known" and the claim that each x_i has an "open back path," the claim of the theorem seems not to be well-formed. The notion of "open back path" is defined in two inequivalent ways. The formal definition is:

> OBP: *x(t)* has an open back path with respect to $x_e(0)$ just in case at any earlier time *t'* there is some cause *u(t')* of *x(t)*, and it is both true and *known to be true* that *u(t')* can cause $x_e(0)$ only by causing *x(t)*.

We will treat her references to what is "known" as extensional specifications of restrictions on a class of possible structures. The alternative structures are essentially alternative directed acyclic graphs, and the question is: under what restrictions on the alternatives can causal properties be reliably identified.

There is still an ambiguity about "*u(t')* can cause $x_e(0)$ only by causing *x(t)*." We take it to mean that every path from *u(t')* to $x_e(0)$ contains *x(t)*. (If in addition it were required that there *exist* such a path, the entire claim would be uninterestingly trivial.) Finally, she says that "each x_i has an open back path" but as defined, having an open back path is relational, and she does not specify with respect to what it is that x_i has an open back path. We take her to mean that each x_i has an open back path with respect to x_e. So a consequence of her claim seems to be this:

> **Claim:** Let x_i, $i = 1, \ldots, m$, be a set of variables and let *G* be an acyclic directed graph having vertices x_i, x_e, and u_k (with *k* ranging over some finite index set) such that each vertex is a

variable that is a linear function of the vertices adjacent to it with edges directed into it, and

(i) there is no directed path from x_e to any x_i;

(ii) $x_e = \Sigma a_i x_i$

(iii) each x_i on the r.h.s. of (ii) that has a directed path to x_e has such a path to x_e that does not contain any x_j on the r.h.s. of (ii) that is not equal to x_j;

(iv) for each x_j having a directed path to x_e there exists a variable u_i having a directed path to x_i and such that every directed path from u_i to x_e contains x_i.

Then for each x_i there is a directed path to x_e.

Cartwright's own putative counterexample (which she rejects) really is a genuine counterexample to this claim: unless Spearman's principle is invoked, it refutes the claim. She starts with a graph (Fig. 5) with the true causal equations:

$$x_1 = u_1$$
$$u_2 = g\,y$$
$$x_2 = a\,x_1 + u_2$$
$$x_e = bx_1 + cy + u_3$$

and considers whether the equation

$$x_e = b'x_2 + u_3$$

with $b' = b/a$ can be derived from them. It can, indeed, provided the numbers a, c, and g are chosen so that $ac = -bg$. For by simple substitutions in the four equations we obtain $x_e = (b/a)x2 - (bg/a)y + cy + u3$. The open back path requirement is met not only with respect to x_e, but with respect to all other variables as well.

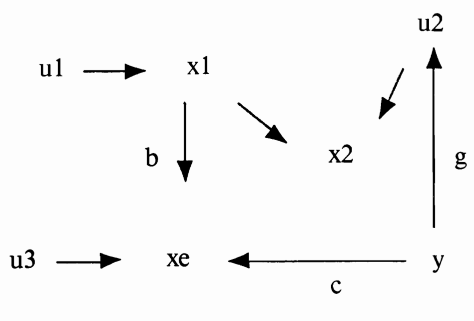

FIGURE 5

NOTES

Research for this report was supported by the Office of Naval Research under Contract Number N00014-88-K-0194. The first author was supported by a fellowship from the John Simon Guggenheim Memorial Foundation and by NSF grant Number DIR-882-1203. We wish to thank Brian Skyrms for helpful comments.

1. More precisely, we assume that if E is the set of exogenous variables in the model, for each x in E, x is independent of $E - \{x\}$.
2. We also assume that each variable in the graph is caused by some "error variable" of non-zero variance which represents all of the unknown causes. The error variables are not explicitly depicted in the graph. Hence, the actual equation for Z should include an error term, and $Z = bY + cW + \epsilon$.
3. The rules for determining which conditional independencies are implied by a given causal structure are described in Section II.
4. The difficult but important problem of determining when we have measured all common causes of variables in S, and what inferences can be made if we have not, are discussed in detail in Spirtes 1990, and Spirtes forthcoming.

REFERENCES

Cartwright, N. *Nature's Capacities and their Measurement.* London: Oxford UP, 1989.

Glymour, C., R. Scheines, P. Spirtes, and K. Kelly. *Discovering Causal Structure: Artificial Intelligence, Philosophy of Science and Statistical Modeling.* New York: Academic Press, 1987.

Kiiveri, H. and T. Speed. "Structural Analysis of Multivariate Data: A Review" in S. Leinhardt, ed., *Sociological Methodology.* San Francisco: Jossey Bass, 1982.

Pearl, J. *Probabilistic Reasoning in Intelligent Systems.* San Mateo, CA: Morgan Kaufmann, 1988.

Spirtes, P., C. Glymour, and R. Scheines. "Causality from Probability" in G. McKee, ed., *Evolving Knowledge in Natural and Artificial Intelligence.* New York: Pitman, 1990, and in *Proceedings of the Conference on Advanced Computing for the Social Sciences,* Williamsburg, VA: 1990.

———. "From Probability to Causality." *Philosophical Studies,* forthcoming.

Spirtes, P., R. Scheines, and C. Glymour. "Simulation Studies of the Reliability of the TETRAD II, LISREL VI and EQS Programs." *Sociological Methods and Research,* forthcoming.

Verma, T. and J. Pearl. "On Equivalence of Causal Models." Technical Report, UCLA Department of Computer Science, April 1990.

Wermuth, N. and S. Lauritzen. "Graphical and Recursive Models for Contingency Tables." *Biometrika, 72* (1983) 537–52.