CLARK GLYMOUR

RABBIT HUNTING

Twenty years ago, Nancy Cartwright wrote a perceptive essay in which she clearly distinguished causal relations from associations, introduced philosophers to Simpson's paradox, articulated the difficulties for reductive probabilistic analyses of causation that flow from these observations, and connected causal relations with strategies of action (Cartwright 1979). Five years later, without appreciating her essay, I and my (then) students began to develop formal representations of causal and probabilistic relations, which, subsequently informed by the work of computer scientists and stattionsssssss-

may similarly imply characteristic sets of tetrad equations. For example, a model in which two measured variables, i and j, share a common unobserved cause, and two other measured variables, and , share a second unobserved cause, implies a single tetrad equation:

$$_{i} \quad _{j} = \quad _{i} \quad _{j}$$

no matter how the unobserved common causes are causally related to one another. Costner and Schoenberg used this observation to propose heuristic methods for modifying initial latent variable models that fail statistical tests on sample data.

The general methodological point of *Discovering Causal Structure* was that social scientific practice is unnecessarily dogmatic, that usually very few alternative explanations are entertained, and that 8(p)med cau-330(p4)81ustetic, t -1.188 TD [(anoth

fee g

partial correlations each model implies. Blalock's procedure was similar to Spearman's and Costner's in that it did not depend on the particular values of the linear coefficients. Blalock carried out his procedure only for simple models with no more than four variables, and he and subsequent social statisticians provided no general algorithm. We provided a general algorithm for computing the first order vanishing partial correlations implied by any recursive linear model without unobserved common causes and with independent errors, a result that seems pitiful in retrospect. Like Blalock, we offered no algorithmic procedure for using these constraints in searching for causal explanations.

The rest of *Discovering Causal Structure* was devoted to justifying heuristic search, explaining the procedures and the methodological intuitions behind them, illustrating their application on well studied sets of social data, and giving proofs. In applications, the search procedures were used to find models with free parameters (the linear coefficients and the variances of unobserved variables, assuming the normal family of distributions). The numerical values for the parameters were then estimated, and the fit models estimated, using a standard statistical package. The illustrative applications typically found plausible, better fitting alternatives to causal models in the social science literature. The single empirically independently verified application of the method was to predict, without prior knowledge, the order in which several questions in a famous sociometric questionnaire had been asked. Perhaps we were lucky.

Nancy Cartwright devoted sixteen pages of *Nature's Capacities and Their Measurement* to criticizing *Discovering Causal Structure.*

> The first and most important difference between my point of view and that argued in *Discovering Causal Structure* has already been registered. I insist that scientific hypotheses be tested. Glymour, Schemes, Kelly and Spirtes despair of ever having enough knowledge to execute a reliable test. (1989, 72)

What Cartwright described as our ill-founded "despair" was our emphasis on this: the fact that a statistical model passes a significance test at some alpha left is insufficient for the truth of the model, since many distinct models may pass the same test, and conventional statistical methodology had no method of finding the alternatives. That is true, and Cartwright said nothing to rebut it.

Next, simplicity.

> They assume that structures that are simple are more likely to be true than ones that are complex. I maintain just the opposite . . . have argued that nature is complex through and through: even at the level of fundamental theory, simplicity is gained only at the cost of misrepresentation . . . Glymour, Schemes, Kelly and Spirtes believe that simpler models are

the principle, some of which she illustrated with a discussion of a case we considered, the Transitional Aid Research Project (TARP).

Separately in Texas and in Georgia, randomly selected groups of newly released fellows were given monthly payments for six months through the respective state unemployment commissions. After a year, rearrest rates for these groups were compared with rearrest rates for felons released at the same time in the respective states. In Texas there was no difference in rearrest rates between the two groups, and, likewise, in Georgia there was no difference in rearrest rates for the two groups. No data were obtained on the actual employment in this period of either the treatment or the control groups. The project leaders concluded, nonetheless, that these facts showed that payments to newly released felons reduce crime. They justified that odd conclusion in this way: in the experimental set-up, payments through the unemployment commission reduced the recipients propensity to work (supposition); unemployment caused the recipients to engage in crime (sociological theory); but since there was no difference in recidivism between the groups that received payments and the groups that did not (empirical data), the payments must have caused a compensating tendency not to do crime (conclusion). The two mechanisms perfectly canceled one another. The explanation is a straightforward violation of Spearman's principle.

In protest to these inferences Hans Zeisel, an eminent sociologist, very publicly resigned from the committee overseeing the experiment. Zeisel thought the straightforward and obvious explanation of the data was that payments (at least at the amounts in the experiments) do not influence recidivism. Our methods agree with Zeisel's in rejecting the arguments and the conclusion of the project leaders, and in thinking the experiment is evidence that payments have no influence on recidivism, but we went on for two pages to dispute Zeisel's claim that randomized experiments always have univocal interpretations, and that the only posw-374(claFsnc)-41n et30(t4T6;tforw

Consider then the probabilities of hypotheses. Although our book did not give a Bayesian analysis, it is straightforward to do so. The probability after the model with free parameters is specified.) With independent errors, that representation is *isomorphic* to the directed graphical representation we use. And, of course, the graphical representation is not new either; it was sixty years old when we used it and perfectly common, just as we represented it, in the social science literature in the thirty years preceding our book.

So what is wrong with our "theory form"?

> The upshot of this implementation of Spearman's Principle is to reduce the information given in a causal theory from that implied by the full set of equations to just what is available from the corresponding causal pictures . . . This move from the old theory form to the new one is total and irreversible in the Glymour, Schemes, Kelly and Spirtes methodology, since the computer program they designed to rank causal theories chooses only among causal structures. It never looks at sets of equations, where numerical values need to be filled in. I think this is a mistake, both for tactical and philosophical reasons (1989, 76).

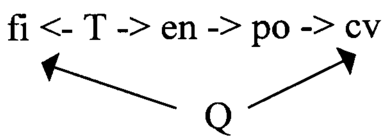
There is nothing "total and irreversible" about the graphical representation that severs it from equational representations with free parameters. Four pages (68–72) of *Discovering Causal Structure* are devoted to describing as well. This makes causal laws fundamentally qualitative: it supposes that in nature only facts about what causes what are important; facts about strength of influences are set by nature at best as an afterthought. I take it, by contrast, that the numbers matter, and that they can be relied on just as much as the presence or absence of the causal relations themselves (1989, 76–7)

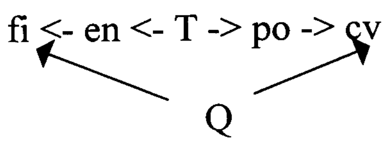
Parallel reasoning to Cartwright's: It makes sense to try to catch rabbits by their ears rather than their tails only if they are more likely to have ears than tails. Our reasons for search over graphical structures rather than systems of equations had nothing to do with whether the existence of causal relations is more real than numerical measures of their strength, whatever that means; the reasons had everything to do with reliability and computational feasibility of search.

The value of searching over graphical structure can perhaps be illustrated by considering the numerically based algorithmic model elaboration procedures standardly used in 1987 (and even now), when *Discovering Causal Structure* was published. Cartwright did not mention them. Their strategy begins with a linear, latent variable model with free parameters and tests the model on sample data at some specified alpha level. In 1987, directed edge that, when added to the initial graph, implied all members of H and a proper subset of I, a model M' containing that edge was created, and its corresponding set I' stored. The procedure then branched over all the elaborated models and repeated the process. A branch of search terminated when no further elaboration could reduce the implied tetrad equations without reducing the implied tetrad equations in H.

The critical difference in the procedures is that, for computational reasons, the conventional search could not afford to branch, and so had to make arbitrary choices, whereas our search, which required no numerical analysis, could and did branch. The difference in search strategies made a considerable difference in reliability. In an enormous simulation study using structures typical in social science models, with randomly assigned parameter values and a variety of sample sizes, the popular beam searches produced the correct answer in 11 to 13% of cases, depending on sample size and the particular algorithm used. Graphical search produced a set could not reconstruct their data from their source. They did not reply to a request from my collaborator, Richard Scheines, for details on how the correlations were obtained.) They simultaneously linearly regressed po on fi, en, and cv, found a positive regression coefficient for fi, and concluded that they had shown that foreign investment caused political exclusion.

Examining their correlations, we found that political exclusion and foreign investment are uncorrelated when energy is controlled for, and that energy and absence of civil liberties are uncorrelated when political exclusion is controlled for. We said these vanishing partial correlations, which are very robust, "are the kind of relationship among correlations that can be explained by causal structure" (1987, 177), we offered some models that explain them in that way. For example, graphically:





where *T* and *Q* are unobserved causes. We did not claim any of these models are true, but did claim they are better explanations of the correlations and time order constraints reported by Timberlake and Williams than is the regression model in which the causal structure is assumed a priori and the vanishing partial correlations are accommodated by the numerical values of the coefficients. Since it involved no automated search, our analysis was exactly the sort that Hubert Blalock could have given.

The logic of Cartwright's discussion is difficult to follow. There is a formal point that may have been what she was after, namely that there exist (normal) probability distributions that do not satisfy Spearman's principle for any directed acyclic graph, with or without latent variables.

I will give her discussion in the sequence she did, changing only notation to agree with mine. First Cartwright asked the reader to assume, contrary to fact, that foreign investment and political exclusion are uncorrelated. Then she asked that the reader assume that two causal claims of the regression model are correct – en is a direct cause of po and cv is a direct cause of po, although there is nothing but the regression model to justify these assumptions. Then she asked the reader to assume that the following three second order partial correlations do not vanish, although she made no showing of this assumption from the data:

fi, po controlling for en and cv
en, po, controlling for fi and cv
cv, po controlling for en and fi

Each of their [i.e., our] structures reverses the causal order of (po) and (cv) (from the order in the TW model) … Since the methods described in Chapter 1 (of *Nature's Capacities and Their Measurement*) assume that temporal order between causes and effects is fixed, a structure in which (fi, cv and en all precede po), as they do in (Timberlake and Williams' model), will serve better for comparing the two approaches (1989).

Her point is that in our models po causes cv, whereas in Timberlake and Williams' regression model the reverse is true. She did not note that they gave no basis for their assumption.

There follows in her book a new graphical model which, she says, implies that po and fi are uncorrelated and also implies the two vanishing partial correlations we found from Timberlake and Williams correlation matrix. The model is:

en po $\leftarrow T \rightarrow$ cv $\leftarrow$ fi

She did not note that, unlike the time order of po and cv about which Timberlake and Williams provided an assumption but no information, they *did* specify that fi is measured at a later time than cv. Unlike our models, Cartwright's model really does violate what is known about the time order.

This structure, she wrote, implies the vanishing correlation of fi and

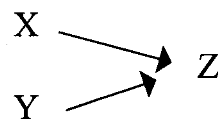
tun(at)-32(sone)-7(s)f5(as3-26G)-28ly-ls, and does not imply

ρ fi,po.en,cv = 0
ρ en,po.fi,cv = 0
ρ cv,po.en,fi = 0

and does not have po cause fi, en or cv,

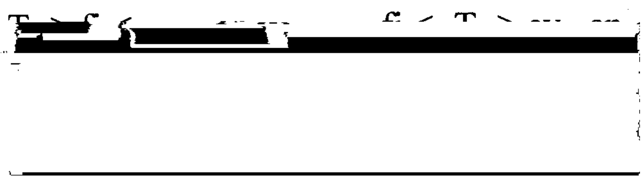
ρ fi causes po.

The assumptions are overkill. In any linear model, a structure of the form



implies that X, Y are correlated controlling for Z. (It does not matter whether the associations between X, Y, and Z are produced by X, Y causing Z or by a unmeasured common causes, or both.) Hence no linear model in which cv and en are each correlated with po, and po is not a cause of either of them, can imply that cv, en.po = 0. But why should it matter to our proposals that a completely imaginary set of constraints should not be explicable by any linear model? Presumably it should matter only if such circumstances are common, and there is some other method for finding the true structure when they arise. Cartwright did nothing either through an empirical survey or through mathematical analysis, to show that such constraints commonly occur. (Six years later we proved that, in the natural measure on the linear coefficients, such constraints have probability zero. See below.)

She continued by offering still another model which she claimed includes our “favoured hypothesis, that foreign investment does not cause repression, and does account for all the data, though of course not on the basis of structure alone” (1989, 83). The point seems to be a charitable effort on her part to formulate a model that saves the data (although it is not clear which data, real or imaginary, she meant to save), incorporates our “favored hypothesis” (although we had no stake in the particular causal claim, only a preference for certain explanatory relations), and corresponds to a time order which she seems to have thought was independently known. She presented the model graphically, as on the left below



and claimed that the data (whichever) cannot distinguish between this structure and Timberlake and Williams' model. Her version of Timberlake and Williams' model – which is not the real one – is shown on the right above. Her version leaves out the correlations of fi and cv with en implicit in the regression model.[1]

“recursive” linear models with independent errors. Then, in 1989, we read Judea Pearl’s book, *Probabilistic Reasoning in Intelligent Systems*, which had appeared the year before, and the lights came on.

In the early 1980s, a number of statisticians had formalized the relation between directed acyclic graphs and the vanishing partial correlations they imply in corresponding linear models with independent errors, and, more generally, between directed acyclic graphs and conditional independence. The crux of the connection was called the (local) Markov condition, and is a generalization of Reichenbach’s notion of screening off. Formally, the Markov condition is simply a restriction on how directed graphs whose vertices are variables are to be paired with probability distributions over the space of possible joint assignments of values to the variables. A pair $\langle G, \mathrm{Pr} \rangle$, G a directed graph and Pr such a probability distribution, satisfies the Markov condition if and only if for each variable X represented by a vertex **X** in G, conditional on the parents of **X** (in G) X is independent of any set of variables, none of whose members are represented by vertices that are descendants of **X** in G. In linear models with normal distributions, conditional independence is vanishing partial correlation, and the Markov condition can also be reformulated for vanishing partial correlations even in non-normally distributed linear models with independent errors.

Pearl not only reviewed this work, he and his students did something of great importance: they used it to provide a fast algorithm to decide, for any directed acyclic graph and any conditional independence statement involving only variables represented by vertices in that graph, whether the Markov condition applied to the graph implies the conditional independence. The algorithm used a graphical property Pearl discovered and called d-separation, although to add confusion it is now sometimes called the (global) Markov property. It is straightforward to prove that the Markov condition is necessarily true of any system of functional dependencies among variables in which the exogenous variables (those of zero in degree in the graph) are independently distributed. So, with d-separation, we could now compute the vanishing partial correlations of any order implied by any directed acyclic graph, and hence by any linear recursive equational model with independent errors. (It later (in 1994) became clear that, in one respect, d-separation is a more generally applicable notion than the ion than v253(gen)-1(han)-263(raph,) the assumption, which we now call faithfulness, that all conditional independence relations in the probability distribution follow from the Markov condition applied to the graph with which the distribution is paired. For linear models, faithfulness is a generalization of Spearman's principle.

In his 1988 book, Pearl explicitly rejected the idea that the graphical structures he described might be used to describe any model-independent causal relations. From our work on linear latent variable models, I and my colleagues had a quite different view, and it proved fruitful. In 1990, Spirtes, Richard Schemes and I used d-separation and tests of conditional independence (or vanishing partial correlations) in an algorithm for constructing causal models from data, provided there are no unrecorded common causes of measured variables, and assuming the Markov condition is true of causal relations and the probabilities of variable values. We also suggested that related searches could be found for latent variable to philosophers, but the book has been the subject of several criticisms, with that. Still they do not want to buy from Cheap-but-Dirty because they object to the pollutants that are emitted as a by-product whenever the chemical is produced.

That is what really is going on, but Cheap-but-Dirty will not admit to it. They suggest that it must be the use of the chemical in the sewage plant itself that produces the pollution.

made about modeling with discrete or continuous variables, data must be differenced to remove auto-correlation, and on and on. The program allows the user to specify a range of assumptions adapted to the “individual circumstances”: latent variables can be allowed or forbidden, and particular causal connections can be forbidden or required.

I will give five examples of positive causal information produced by the procedures, cases where, either by independent interventions or by well established independent knowledge not used in the data analysis, predictions of the procedure were established.

Case 1.

miscalibrations resulting from the space environment. Using our program, physicists at the Swedish Institute for Space Physics concluded that the instrument reliably measured total concentrations of heavy ions and total concentrations of light ions, but not concentrations of particular species. After recalibration of the data interpretation software, the differences from theory were reduced by half automated procedure performed comparably to the human expert, in some respects slightly worse, in other respects slightly better. This is a problem "hypothetico-deductive", which is in some sense true, but chiefly indicates the poverty of philosophical vocabulary in talking about search. I do not have enough of her forthcoming book to know what methods, if any, it advocates, but the chapter I have read suggests methods of inquiry must now be "hypothetico deductive". But hypothetico-deductive method is not a *method* of inquiry; it is at most a cog in a method, and as conventionally used in social statistics, where a few guesses are tested and all other possible guesses ignored, not even that.

The majority of *Nature's Capacities and Their Laws* is really devoted to developing a metaphysical conception of probabilistic causation that I think is perceptive and has proved fruitful. I have not described Cartwright's positive metaphysic because, no matter the offence it may give, this essay is defensive. But I wish to praise her metaphysic. By my lights, philosophy of science should be largely judged by its contribution to scientific progress, and by that measure *Nature's Capacities and Their Laws* stands out. Patricia Cheng recently proposed a psychological model of human judgement, justified by an impressive array of experimental results. The picture of causation Cheng employs is Cartwright's, and although Cartwright is not cited, Cheng tells me she had read

Intelligence, and Cheng, P. W.: 1997, 'From Covariation to Causation: A Causal Power Theory', *Psychological Review* **104**, 367–405.

REFERENCES

Cartwright, N.: 1979, 'Causal Laws and Effective Strategies', *Nous* **13**, Reprinted in Cartwright (1983).

Cartwright, N.: 1983, *How the Laws of Physics Lie*, Oxford University Press, Oxford.

Cartwright, N.: 1989, *Nature's Capacities and Their Measurement*, Oxford University