The Monist, 2022, 00, 1–29
https://doi.org/10.1093/monist/onab034
Article

The Statistical Nature of Causation

David Papineau*

ABSTRACT

Causation is a macroscopic phenomenon. The temporal asymmetry displayed by causation has no counterpart in the fundamental dynamics of the microscopic world. So causation must somehow emerge from the underlying dynamics, along with other macroscopic phenomena like entropy increase and the arrow of radiation. In this paper I shall offer an account of causation that accounts for this emergence. I shall show how macroscopic events fall into lawlike patterns that are subject to the random influences of microscopic processes. It is the probabilistic independence of these random influences that constitutes these laws as temporally asymmetric and causal. I shall approach these issues by considering 'causal inference' techniques that allow causal relations to be inferred from sets of observed correlations. I shall show that these techniques are best explained by a reduction of causation to structures of equations with probabilistically independent exogenous terms. This exogenous probabilistic independence imposes a recursive order on these equations and a consequent distinction between dependent and independent variables that lines up with the temporal asymmetry of causation.

1. INTRODUCTION

Causation is a macroscopic phenomenon. The temporal asymmetry displayed by causation has no counterpart in the fundamental dynamics of the microscopic world. So causation must somehow emerge from the underlying dynamics, along with other macroscopic phenomena like entropy increase and the arrow of radiation.

In this paper I shall offer an account of causation that accounts for this emergence. I shall show how macroscopic events fall into lawlike patterns that are subject to the random influences of microscopic processes. It is the probabilistic independence of these random influences that constitutes these laws as temporally asymmetric and causal.

I shall approach these issues by considering "causal inference" techniques that allow causal relations to be inferred from sets of observed correlations. I shall show that these techniques are best explained by a reduction of causation to structures of equations with probabilistically independent exogenous terms. This exogenous probabilistic independence imposes a recursive order on these equations and a

*King's College London, UK

consequent distinction between dependent and independent variables that lines up with the temporal asymmetry of causation.[1]

Philosophers in the tradition of David Lewis have sought in various ways to account for the temporal asymmetry of causation in terms of the "asymmetry of overdetermination."[2] These accounts are consonant with the analysis that I shall offer, but from my perspective they start the story too late. The "asymmetry of overdetermination" is not something to be assumed, but a macroscopic phenomenon that itself demands explanation. My analysis will have the virtue of displaying the Lewisian asymmetry as a consequence of the asymmetric nature of causation.

2. EXPLAINING CAUSAL INFERENCE

For over a hundred years nonexperimental scientists have been inferring asymmetrical causal conclusions from correlational data. Curiously, none of the main philosophical theories of causation cast any light on why these techniques work. What in the nature of causation allows such inferences to proceed? As far as I know, nobody working on counterfactual, or regularity, or process, or dispositional theories of causation so much as asks this question.

The only thinkers who have addressed this issue are those philosophers in the older minority tradition of "probabilistic theories" of causation. These theories attempt to explain the inference techniques by reducing causal relationships directly to correlational ones. In the middle of the last century Hans Reichenbach (1956), I.J. Good (1961-62) and Patrick Suppes (1970) all offered variations on this theme, and more recently Wolfgang Spohn (2001), Clark Glymour (2004), Gerhard Schurz and Alexander Gebharter (2016), and David Papineau (1992, 2001) have drawn on the analysis of "Bayesian networks" to develop more sophisticated versions of this strategy. However, as I shall show, this tradition ties causation too closely to correlations. Because of this, it cannot cope with "faithfulness failures" where correlations are misleading about causal links, nor is it able explain the relationship between probabilistic causal connections and single-case actual causation.

My strategy in this paper will be to offer a different reductive analysis. The idea goes back to H.A. Simon and others in the 1950s and 60s (Simon 1953; Blalock 1964). It seeks to reduce causation to underlying structural equations with probabilistic independent exogenous terms rather than directly to surface correlations. While this idea is often enough aired by practising nonexperimental scientists, it has been largely ignored by philosophers (though see Cartwright 1989 and Hausman 1998). I shall show that this approach holds the key both to the success of the correlational inferential techniques and to the way that causation is a temporally asymmetric macroscopic phenomenon.

Much recent work on causation characterises itself as adopting an "interventionist" approach. This term covers a number of different ideas, and their detailed relation to my own analysis will have to be left to further work. By and large, though, the general interventionist programme is consonant with my approach. My attitude to this programme is not that it is mistaken, but that it does not go deep enough.

In effect, I see the interventionist programme as falling between the two stools of (a) leaving the causal inference techniques unexplained and (b) reducing causes directly to correlations in the style of probabilistic theories. Thus James Woodward in his *Making Things Happen* (2003, 51) specifies that X is a (total) cause of Y just in case the probability of Y would change if X were changed *by an intervention*—where an intervention is defined as a way of altering X that is statistically independent of the other *causes* of Y. Now, by so equating causation with correlation under intervention, Woodward's formula does forge some link between correlations and causes, and thereby casts some light on the causal inference techniques (Hausman and Woodward 1999). But it cannot fully explain how causes can be inferred from correlations alone, since it appeals circularly to causal conditions in defining an "intervention." At the same time, precisely because it ties causation directly to correlation, it is doubtful whether Woodward's approach copes with faithfulness failures any better than the reductive probabilistic theories (Strevens 2007, 2008; Woodward 2008). My intention in this paper is to do better—not by dismissing the interventionist approach, but by offering a fuller analysis that can account for its successes.

3. BRIDGE PRINCIPLES

Let me quickly illustrate the kind of causal inference techniques at issue. Suppose that good examination results (E) are positively correlated[3] with attendance at private rather than state schools (S), but this correlation disappears when we control for parental income (P)—the correlation is *screened off* by P, as it is said, in that E is no longer more likely given S once we hold P fixed. We conclude that, despite the prima facie indication that private schools causally influence examination results, this correlation is in fact "spurious" in the sense that schools S don't affect examination results E after all; rather both are effects of the common cause parental income, as in the following causal *directed acyclic structure* ("DAS" henceforth):[4]

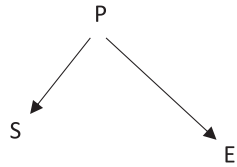

DIAGRAM 1

Correlational-to-causal inferences like these are natural enough, but they do not wear their rationale on their sleeve. Recent work in the tradition of "Bayesian networks" has regimented the principles behind them. (See, for example, Spirtes et al. 1993, Pearl 2000, Peters et al. 2017.) In this section and the next, I shall articulate these principles and explain their inferential power, taking their acceptability as given. Once we are clear about how they work, we can then turn to questions about their truth and metaphysical status.

It will be convenient for what follows to say that two variables X and Y are *causally linked* if X causes Y (possibly indirectly via intermediaries), or Y causes X (again

possibly indirectly), or X and Y have a (possibly indirect) common cause—but not if X and Y only have a common effect.

One issue that arises at this point is what is meant by saying that one generic variable "causes" another, in the sense in which, say, parental income might cause examination results or smoking might cause lung cancer. Presumably such claims are some kind of generalization over particular cases in which definite values of these variables bear some more specific relation to each other. But how to fill this out is by no means obvious. Still, it will be convenient for our purposes to take the generic notion of cause as read for the time being. By the end of the paper I will be able to explain it.

Inferring causes from correlations hinges on two kinds of principles—I shall call them "bridge principles" henceforth. On the one hand are a pair of principles licensing moves *from* correlations *to* causes, and on the other a pair licensing moves *from absence* of correlation *to absence* of causes. (The former pair are often presented together as the "Causal Markov Condition" and the latter pair together as the "Faithfulness Condition." But it will be more illuminating to unpack them as follows.)

Let us start with the pair of correlation-to-cause principles. First and simplest is what I shall call the *Linkage Principle*:

(1) If two variables are correlated, then they must be causally linked.

And to this can be added a *Conditional Linkage Principle*:

(2) If two correlated variables remain conditionally correlated after we control for other variables {X}, then they must be casually linked by one or more paths that do not go via {X}.

These two principles specify that correlations always indicate a causal link: correlated variables must either be related as cause and effect or they must have a common cause. Moreover, correlations that persist even after controlling for other variables indicate casual links that bypass those controlling variables.[5]

Now for the converse no-correlation-to-no-cause principles.

The *Unlinkage Principle* says:

(3) If two variables are uncorrelated, then they are not causally linked.

And a further *Conditional Unlinkage Principle* says:

(4) If two correlated variables are screened off by other variables {X}, then they are not causally linked by any chains of variables that do not contain any of {X}.

These two principles now tell us that variables that are *not* correlated are *not* causally linked: they can't cause each other or have a common cause. Moreover, two variables that cease to be correlated when we control for other variables cannot be linked in any ways that bypass those controlling variables.[6]

Together these principles suffice to fix a causal order among variables displaying a sufficiently rich set of correlations.

Of course, they don't fix a causal order among any set of correlated variables. To revert to our original schools (S)-examinations (E)-parental income (P) example, the pairwise correlations between all three variables, plus the screening-off of the S-E correlation by P, is in itself consistent, not only with the possibility suggested earlier, where P causes both S and E, as in Diagram 1, but also with

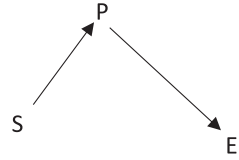

DIAGRAM 2

and

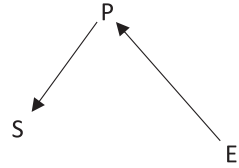

DIAGRAM 3

Still, this indeterminacy would be resolved if we were given some further variable A that were correlated solely with S and another B that were correlated solely with E. Then the only causal structure consistent with the bridge principles would be

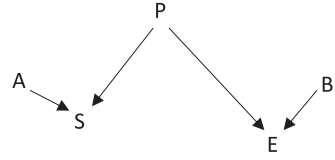

DIAGRAM 4

This illustrates a result that can be proved in full generality. Whenever the correlations between some set of variables do not allow the bridge principles to fix their causal relationships uniquely, there will always be possible correlations involving further possible variables that will so suffice. (Theorem 4.6, Spirtes et al. 1993, 94.)

Let me observe at this point that the bridge principles fix causal order, when they do, without resorting to any information about the temporal ordering of the relevant variables. Yet we can expect that, when they do so fix causal order, the variables identified as causes will in reality always precede their effects in time. If this is so, this must be because this temporal ordering of variables is implicit in the empirical correlations displayed by causally related sets of variables. The arrangement of correlations is itself

asymmetrically distributed in time. This augurs well for the project of understanding how the temporal asymmetry of causation can emerge in a world with a temporally symmetric fundamental dynamics.

Of course, empirical researchers don't always need to infer their causal conclusions from correlations alone. In practice they will standardly narrow down the causal possibilities and simplify their inferential task by helping themselves to prior causal knowledge, courtesy of common sense or the temporal ordering of variables. So for example, in our initial example, they would quite sensibly have taken it as given that temporally later examinations E results cannot cause earlier school type S or earlier parental income P. However, as we have seen, this kind of assistance from common sense or temporal ordering is by no means essential.

Philosophers sometimes emphasize how particular sets of correlations can leave causal structure undetermined even given the bridge laws, as did the original S-P-E correlations in our example, and how empirical researchers will standardly invoke prior causal knowledge to resolve the indeterminacy.[7] These points are of course true, but they should not be allowed to obscure the mathematical fact that in such cases richer sets of possible correlations would always suffice to fix causal order on their own. (Whether reality will always provide such richer sets of correlation is of course a further question, to be decided by the empirical facts rather than mathematical proof. We shall return to this issue at various points below.)

4. CAUSAL SUFFICIENCY

The points made so far might suggest the reductive idea that there is nothing more to causal relations than the patterns of correlation from which they can be inferred via the bridge principles. In effect, this would be to view the bridge principles as necessary truths that encode the way that causal relations are implicit in correlational structures.

I shall explore this idea further in the next section. But first an immediate issue must be addressed. I have shown that the bridge principles can determine a causal order given the correlations *among a given set of variables*. But this leaves it open that a causal order so determined might be overturned if the set of variables were expanded. This possibility clearly threatens the idea that causal relations are nothing over and above the correlational patterns that imply them. After all, our aim is to analyse the nature of *real* causal relations, not of *apparent* causal relations *relative to* an arbitrary selection of variables.

This worry is by no means an idle one. Suppose that, in the way described in the last section, the correlations among some {A, B, S, P, E} determine:

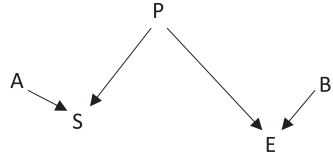

DIAGRAM 4

Despite this unique determination of a causal order by the correlations, it remains perfectly possible that in reality P does not cause S, but rather both are effects of some further G. (Imagine, not entirely implausibly, that parental income P per se has no effect on school type S, but that both are effects of grandparental income G.) In that case, the correlations among {A, B, S, P, E} would remain just as observed, but the bridge-principle driven conclusion that P causes S would simply be wrong, and the true causal structure would be

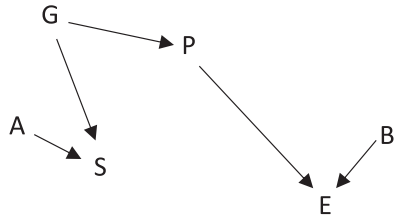

DIAGRAM 5

However, while this is a real danger, it is significantly limited. To see why, note that bringing in extra variables will not itself alter the correlations among the variables in some original set. For example, expanding the analysis by including G won't stop P being systematically correlated with S. At most, extra variables will screen off correlations that weren't screened off in the original variable set—with the result that, if we now apply the bridge principles to the expanded set of variables, any such newly exposed screening off will indicate that causal links presented as direct by the original set are in fact only indirect causal links.

Now, indirect causal links are of two kinds—either one variable causes another via an intermediary, or two variables have a common cause. In the first kind of case, expanding our variable set will not really have overturned any causal conclusion, since it will only have shown that some causal link proceeds via intermediaries, as would in any case have been assumed. So it is only the second kind of case, where the extra variable turns out to be a common cause of two variables in the original set, that the casual verdicts delivered by the original correlations will be reversed.

This now shows that there will be no overturning of verdicts delivered by the bridge principles as long as we have a *causally sufficient* set of variables, in the sense of a set that does not omit any variables that turn out on expansion to be common causes of variables included in the original set. And this now opens the way to the reductive project once more. A revised reductive suggestion would now be that causal relations are nothing over and above those patterns of correlation that imply them, courtesy of the bridge principles, in any *causally sufficient* set of variables. (Would not the need to specify *causal* sufficiency here render this suggestion inadmissibly circular as a reduction of causation? But this specification can be finessed away. We can simply say causal relations are nothing over and above the patterns of correlation that imply them in sets of variables whose verdicts are not overturned by the inclusion of further variables.)

5. PROBABILISTIC THEORIES OF CAUSATION

The original probabilistic theories of causation, proposed in the middle of the twentieth century by Reichenbach (1956), Good (1961-63) and Suppes (1970), were all versions on this theme:

> (5) An earlier X causes a later Y if and only if they are positively correlated and this correlation is not screened off by any yet earlier Z.

But there are obvious drawbacks to this formulation. For a start, it appeals to temporal order in its analysis of causal order and thus abandons the search for an independent explanation of why causal relations are asymmetric in time. Moreover, this formulation is ill-suited to accommodate various complex causal structures, as when there are two common causes of two correlated effect variables, with neither cause therefore fully screening the correlation among the effects.

The regimentation of the bridge principles in the Bayesian network tradition allows probabilistic theories of causation to bypass these two difficulties. As we have seen, the bridge principles are capable of determining a causal order among any complex set of correlated variables, and they do so without assuming any information about the temporal ordering of those variables. This then opens the way for reductive theories of the kind suggested above, according to which there is nothing more to causal relations than the patterns of correlation from which they can be inferred via the bridge principles. (See, for example, Spohn 2001, Glymour 2004, Schurz and Gebharter 2016 and Papineau 1992, 2001.)

Can this implicit reduction be transformed into an explicit analysis of causation? Most of the writers just cited do not attempt this, but the theory offered by Daniel Hausman in his *Causal Asymmetries* (1998) can be adapted for this purpose. Hausman himself does not propose an explicit definition of causation in terms of correlations because of the "failures of faithfulness" that I shall address in the next section, but if we put that issue to one side for the moment, we can adapt his analysis and say:

> (6) X causes Y if and only if X is correlated with Y and everything correlated with X is correlated with Y and something correlated with Y is not correlated with X.

The basic idea behind this reduction is that the effects in correlated cause-effect pairs are distinguished from the causes by having probabilistically independent sources of variation, and correlated joint effects of common causes are distinguished from cause-effect pairs by *both* having independent sources of variation.

If we make the assumption that effects do always have such independent sources of variation, then the reductive claim (6) follows from the bridge principles (1)–(4). The need to add this assumption of independent sources of variation to the bridge principles is a reflection of the point, made in the section before last, that while not every set of correlations itself suffices for the bridge principles to determine a causal order, there is always a possible expansion of that set of correlations that will suffice for this. As before, it is an empirical question whether reality will always provide such independent

variation, not something that can be established by metaphysical analysis. We shall come back to this issue in my final section.

6. FAILURES OF FAITHFULNESS

Unfortunately, we cannot rest with this kind of probabilistic reduction of causation. It focuses on the surface symptoms of causation, rather than its underlying nature.

One way to see this is to note that the causal claims analysed by this reduction will leave us without answers to various questions of single-case causation. Suppose, as is no doubt the case, that the bridge principles in conjunction with actual empirical correlations establish that *smoking causes lung cancer*. Now suppose that Joe Bloggs smokes and gets lung cancer. Did his smoking cause his lung cancer? It depends. Even if it is true that "smoking causes lung cancer" in the sense inferred from the bridge principles and the correlations, Joe's genetic make-up might prevent cigarettes from harming him and he might have acquired his lung cancer from asbestos exposure instead.

This shows that there must be more structure to causation than is captured by the kind of causal claims analysed by the proposed probabilistic reduction. I shall be proposing an account of this extra structure in what follows.

Associated with this point are doubts about the status of the bridge principles themselves. The probabilistic reduction argues that they are necessities whose truth falls out of the nature of causation. But it is not obvious that the bridge principles are all true, let alone necessarily true.

All the bridge principles have been questioned in the literature. As it happens, it is my view that the objections to the Causal Markov Condition are misplaced, and that the truth of this condition does indeed fall out of the underlying nature of causation (as I shall argue in section 9 below). But the Faithfulness Condition has a far less secure status. Actual exceptions might be rare, for reasons shortly to be explained, but even if it is generally true is by no means metaphysically guaranteed.

Let me first briefly indicate why I view objections to the Causal Markov Condition as misplaced. Both of its corollaries, the Linkage and Conditional Linkage Principle, have been queried in the literature.

The standard counterexamples to the Linkage Principle are time-series correlations which point to no causal linkage, like the matched values of London bread prices and Venice water levels over the years. As a range of writers have pointed out, however, the Linkage Principle can be qualified so as to exclude correlations with this type of nonstandard time-series construction (Sober 2001; Hoover 2003; Zhang and Spirtes 2014).

Then there are putative counterexamples to the Conditional Linkage Principle. These are cases of correlations which supposedly fail to be screened off when causal intermediaries are controlled for. Wesley Salmon's pool balls (1984) and Nancy Cartwright's polluting factory (2002) are well-known examples. Here the standard response is that the cases omit a full specification of the intermediaries and that that the correlations would disappear if this were rectified.

A different kind of counterexample to the Linkage and Conditional Linkage Principles involves nonlocal quantum correlations like those between measurements on spacelike separated entangled particles ("EPR" correlations henceforth, after

Einstein, Podolsky and Rosen 1935). I shall postpone discussion of such quantum correlations until section 13 below.

Now for the Faithfulness Condition. The problem cases here involve a cause influencing an effect via two separate pathways, one producing the effect and the other preventing it, with the two influences cancelling out, and so leaving us with a zero correlation between the cause and the effect. The classical example, due to Hesslow (1976), involves birth control pills B exerting a direct positive influence on the formation of thromboses T, but also working to prevent T by preventing pregnancies P which themselves operate as a positive cause of thromboses. If these two paths of influence cancel out precisely, then B and T will end up quite uncorrelated, with thromboses just as common among those taking birth control pills as those who don't. (This would be a counterexample to the simple Unlinkage Principle. But similar counterexamples to Conditional Unlinkage can easily be constructed—for instance, just imagine that some C is a common cause of both B and T.)

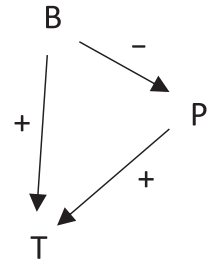

DIAGRAM 6

Cases like these would be clear violations of the Faithfulness Condition. B and T are uncorrelated, even though causally linked. If we applied the bridge principles to this case, we would conclude that the uncorrelated B and T are independent causes of pregnancies P, since they are both correlated with it—which by hypothesis is not the real causal structure.

Now such "failures of faithfulness" are clearly highly unlikely, given the way they depend on an exact cancelling out of influences. Still, this is little consolation to those defending a probabilistic reduction of causation. For that requires the Faithfulness Condition to be a metaphysically necessary consequence of the nature of causation. And cases like Hesslow's, however unlikely they might be, just don't seem metaphysically impossible. It certainly doesn't look as if there is anything in the nature of causation to stop two causal paths cancelling out exactly and leaving us with a null association between cause and effect. I conclude that we need to probe deeper to understand the connection between the bridge principle and the underlying nature of causation. (Before moving on, it is worth observing that, while exact failures of faithfulness might be highly unlikely, and so only of abstract philosophical significance, *approximate* failures of faithfulness are common enough in the real world, and so a practical concern to those nonexperimental scientists who need to start their investigations by using finite sample data to estimate correlations.)

7. STRUCTURAL EQUATIONS

To get to the bottom of the connection between causation and correlation, we need to switch to a different approach, namely the "structural equations" tradition that played a central role in econometrics and sociometrics in the middle of the last century. To revert to our earlier example, this tradition would have dealt with the covariation of schools (S), examination results (E), and parental income (P) by positing a set of linear equations like this:[8]

$$P = e_P \tag{7.1}$$
$$S = aP + e_S \tag{7.2}$$
$$E = bP + cS + e_E \tag{7.3}$$

In such a set of equations, the e-terms that appear only on the right-hand sides are called "exogenous variables" and represent further influences beyond those explicitly included in the study. The other variables on the right-hand side of each equation are called its "independent variables," and the variable on the left-hand side its "dependent variable."

A set of such equations is *recursive* if it can be ordered so that no term appears on the right-hand side as an independent variable unless it has appeared in a previous equation as a dependent variable. In such a case, the equation set will have a directed acyclic structure. In our example, this would be:

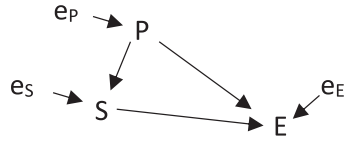

DIAGRAM 7

The regression coefficients a, b, c attaching to the independent variables measure the extent to which the dependent variables vary specifically in response to changes in those independent variables. They capture how much, if at all, the dependent variable "wiggles" when a given independent variable "wiggles" and the other independent variables are held constant.

In our example, we supposed that examination results E don't covary at all with schooling S once parental income P is held constant. So then the regression coefficient c will be zero, and the equations will have the simpler structure:

$$P = e_P \tag{8.1}$$
$$S = aP + e_S \tag{8.2}$$
$$E = bP + e_E \tag{8.3}$$

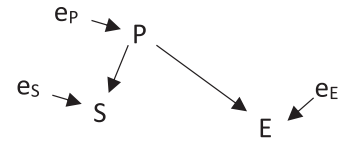

DIAGRAM 8

I have now started representing sets of equations like (7) and (8) by directed acyclic structures—DASs. But note that this is a different kind of DAS from those considered earlier. Where the earlier DASs represented *causal* structures, these new DASs simply represent the order in which variables appear in structures of equations like (7) and (8).

It is both normal and natural, however, to interpret these new equational structures causally, and assume that one variable is a cause of another just in case it is an ancestor of it in the structure of equations—that is, to take our new equation-DASs also to be causal-DASs.

Still, it is not clear that anything said so far *justifies* such an interpretation. After all, if the equations are just *equations*, what is to stop us rearranging terms so that the dependent variables become independent and vice versa? For example, what is to stop us rewriting the equations (8) as follows?

$$S = e*_S \tag{9.1}$$

$$P = \frac{1}{a}S + e*_P \tag{9.2}$$

$$E = bP + \ e_E \tag{9.3}$$

This would then give us this directed acyclic equation structure:

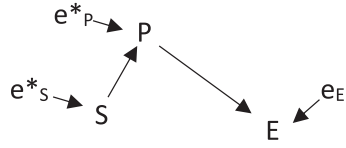

DIAGRAM 9

And, if we were to interpret this structure causally, it would now present S as a cause of P, and P as a cause of E, and S as having no direct causal influence on E except via P.

8. INDEPENDENT EXOGENOUS TERMS

Even so, an approach to causation in terms of structural equations has the resources to meet this challenge. The key idea is that the ordering of variables in a set of equations can capture causal structure only if *the exogenous variables are probabilistically independent*. This requirement promises to decide between the alternative causal hypotheses suggested by equations (8) and (9). If P causes both S and E, then the exogenous terms in the former but not latter equations will be independent, whereas if S causes P which causes E, then the reverse will be true.

This idea was commonplace among econometricians and sociometricians in the middle of the last century. In this connection, note how the requirement of exogenous independence is built into the use of structural equations as a tool for prediction and explanation. Thus consider once more the equations which present school type S as a function of parental income P:

$$P = e_P \tag{7.1}$$
$$S = aP + e_S \tag{7.2}$$

Given some specific value P_k for P, we would naturally use these eqautions to infer that expected school funding is aP_k. But note how this inference hinges on the implicit premise that the extra variation in S is independent of what value of P we have. That is why we can estimate S on the basis of knowing P even while being quite ignorant of the value of e_s.

Observe how this kind of inference doesn't work the other way around. Consider, instead of (7.1–7.2), the rearranged

$$S = e*_S \tag{9.1}$$
$$P = \frac{1}{a}S + e*_P \tag{9.2}$$

If e_s was independent of P in the original (7.1–7.2), then e^*_P won't be independent of S in (9.1–9.2)—for the rearrangement implies that $e^*_P = -e_S/a$—and so now we can't infer that $\frac{1}{a}S$ will be the average value of P given some value of S. The way P varies around $\frac{1}{a}S$ will be different for different values of S, and will depend on how P itself is distributed. (For example, if the median of P is below the average, as we would expect for parental income, then the expected value of P for positive S will generally be less than $\frac{1}{a}S$.)

In my view, the probabilistic independence of exogenous terms in recursive systems of structural equations holds the key to causal direction. This kind of independence means that the values of the dependent variables are due to influences that can be factorised into independent sources. The probabilistic independence of the exogenous terms thus displays them as causally unlinked, while the dependence of the other variables on factorizable influences manifests their status as effects.

So far I have illustrated the idea in a maximally simple case with just two nonexogenous variables. But the idea that causal structure requires independent exogenous terms goes over to more complex cases. To illustrate, let us imagine, contrary to our supposition so far, that schools S do after all exert an extra influence on examination results E, in addition to any direct influence from parental income P. The relevant equations would then be the earlier (7):

$$P = e_P \tag{7.1}$$
$$S = aP + e_S \tag{7.2}$$
$$E = bP + cS + e_E \tag{7.3}$$

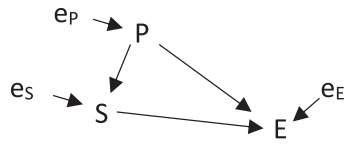

DIAGRAM 7

If the sequence of exogenous terms is independent, then the values of P are fixed by one set of factors e_P, the values of S are fixed by P plus another probabilistically independent set of factors e_S, manifesting the way S is an effect of P and e_S—and finally the values of E are fixed by the values of P and S (which are themselves now correlated) and by yet another set of factors e_E which are probabilistically independent of both P and S. This last independence thus displays E as an effect of all of P, S and e_E.

The underlying idea, then, is that every dependent variable will have an associated exogenous variable that is independent of the other variables it is dependent on. This reflects the assumption, presupposed by the earlier analysis of causation (6) derived from Daniel Hausman's work, that effects will always have sources of variation that are independent of their other causes. But now we have built this assumption into a more structured framework that will prove better suited to deal with failures of faithfulness and with single-case causation.

I have been using linear regression analysis to illustrate the idea that causal structure might depend on independent exogenous terms. But the idea can happily be generalised to other structures of deterministic equations. We needn't restrict ourselves to linear equations, nor to real-valued variables.

Suppose we have any set of variables $X_1, \ldots X_n$ and exogenous terms, $E_1, \ldots E_n$, possibly with values that might be dichotomous, or determinable, as well as quantitative in some way; and suppose we have a set of recursive deterministic equations over these variables of the form

$$X_i = F(X_1, \ldots X_{i-1}, E_i) \tag{10}$$

Then in general, I say, it will be a requirement on these equations capturing causal structure that the exogenous terms be probabilistically independent.

So I now propose the following requirement on causal structures:

(11) X causes Y only if it is an ancestor of Y in a recursive structure of deterministic equations with independent exogenous terms.

In this section I have presupposed that all the variables within any causal structure will be connected by deterministic equations. Quantum mechanics gives us reason to doubt that this is true. In section 13 below I shall modify my analysis to accommodate quantum indeterminism within causal structures. But for the moment it will be useful to continue with assumption of determinism.

9. THE BRIDGE PRINCIPLES RECOVERED

The proposed connection between causation and structural equations casts a new light on the bridge principles that underlie inferences from correlations to causation. Instead of seeing the correlations as providing the substance of causation, as on probabilistic theories of causation, we can now view them as offering indirect evidence for the way variables feature in systems of causally adequate structural equations. On this account, when we use the bridge principles to infer from correlations that X

causes Y, we are in fact inferring that X is an ancestor of Y in a system of causally adequate structural equations.

Crucial in this connection is a mathematical theorem that I shall call the call the "*Determinism-Independence-Markov Result.*" Suppose as before that we have a system of dependent variables $X_1, \ldots X_n$, with probabilistically independent exogenous variables, $E_1, \ldots E_n$, and recursive deterministic equations over these variables of the form $X_i = F(X_1, \ldots X_{i-1}, E_i)$. Then the system will display this Markov property:

> (12) Any variable will be probabilistically independent of every other variable (apart from its descendants) conditional on its parents (where "parent" and "descendant" signify relatedness in the DAS of the relevant equations). (Pearl 2000, Theorem 1.4.1.)

This result is obvious enough. Any two dependent variables that owe their values to disjoint sets of exogenous variables will inherit their independence from the independence of those exogenous variables. Putting it the other way around, two dependent variables will be correlated only if in the equations one descends from the other or they have a common ancestor. Moreover, if a variable does so descend from another, or shares a common ancestor with it, then any correlation between them will disappear if we hold fixed its parents, because any residual variation in the two variables will then again derive from disjoint sets of independent exogenous variables.

Note that as it stands this Determinism-Independence-Markov Result says nothing about *causes* as such. It is a straightforward mathematical claim about the joint probability distribution imposed on all the variables in a system of deterministic equations by the requirement that the exogenous terms be independent. Still, when we combine this theorem with the requirement (11) that causation implies recursive equations with exogenous independence, then the theorem does imply the *Causal* Markov Condition—any *causal* structure will satisfy the Markov property—and therewith the Linkage Principle—correlated variables in a causal structure must be causally linked—and Conditional Linkage Principle—if correlated variables in a causal structure remain correlated when we control for some further variable, then they must be linked by a route that does not involve that further variable.

So the condition on causation (11) proposed in the last section can account for the use of the Linkage and Conditional Linkage Principles to infer causal conclusions from correlational premises. It is noteworthy, though, that this analysis does not simply *posit* that causal structures will satisfy these principles, as did the probabilistic theories of causation considered earlier. Rather it *derives* this from the proposed connection between causation and systems of recursive equations with independent exogenous variables.

It is also noteworthy that the Faithfulness Condition does *not* follow from condition (11). That is just as it should be. As we saw earlier, it is highly implausible to suppose that the Faithfulness Condition is built into the metaphysical nature of causation. True, we can generally expect variables that are linked in a system of equations with exogenous independence to be correlated. The equational links plus the background independence will generally lead to the variables co-varying. In certain

special cases, however, a specific cancelling-out of coefficients will mean that equationally linked variables display overall no correlation.

Recall Hesslow's example in which birth control pills affect thromboses both directly but also via preventing pregnancies which themselves conduce to thromboses. Within the structural-equations framework, this set-up might be realised by the following equations:

$$B = e_B \tag{13.1}$$

$$P = aB + e_P \tag{13.2}$$

$$T = bB + cP + e_T \tag{13.3}$$

Now in this case, and indeed in all cases with this equational structure, there will be no correlation between B and T if the coefficients cancel out exactly and $a + bc = 0$.

From the perspective I have now adopted, however, this kind of case is no longer a problem. I am now assuming causal structures depend on systems of equations with exogenous independence. And Hesslow's example poses no challenge to this assumption. The above equations portray the way that B is a causal ancestor of T via two separate paths. That the coefficients conspire to stop this fact displaying itself in a correlation, as would normally happen, does nothing to undermine the claim that B is doubly a causal ancestor of T. It just shows that the Faithfulness Condition is only a reliable rule of thumb, and not a necessary truth.

So the Faithfulness Condition now falls into its rightful place, as something that empirical researchers can generally rely on, but is in principle open to exceptions. There is nothing in the nature of causation to guarantee that probabilistic independencies should not arise by a cancelling out of parameters. This would be a freakish chance, but it is not built into the nature of causation. Unlike the Causal Markov Condition that takes us from correlations to causal links, the converse Faithfulness Condition that says that causal links display themselves in correlations is only delivered as a reliable rule of thumb.

10. A REDUCTION OF CAUSATION

In the last two sections I have argued that causally connected variables are related by systems of deterministic equations with probabilistically independent exogenous variables, and I have used this to clarify the status of the Linkage and Unlinkage Principles.

But does the connection work the other way around? If the covariation of some variables can be captured by recursive deterministic equations with probabilistically independent exogenous variables, does this imply that the causal ordering of those variables must match this equational ordering?

If this were so, then we could uphold the following reductive analysis of causation:

(14) X causes Y *if* and only if it is an ancestor of Y in a recursive structure of deterministic equations with independent exogenous terms.

However this simple reduction will not work. There are systems of recursive equations with exogenous independence that do not reflect causal structure. If we are to develop an explicit reduction of causation, we will need to take account of these and show how to put them to one side. In this section I shall briefly indicate how this might be done.

Failure of faithfulness yields one kind of case in which equations with exogenous independence do not match causal structure. Consider Hesslow's example once more. The cancelling-out involved means that we end up with birth control pills B and thromboses P being probabilistically independent. So, in addition to the equations (13) representing the real causal structure, we will also have this system of equations with exogenous independence:

$$B = e_B \tag{15.1}$$

$$T = e*_T \tag{15.2}$$

And this would then discredit the proposed reduction of causation (14), since it would imply that B and T must be causally unlinked, which by hypothesis is false.

We also find violations of (14) with certain joint probability distributions involving limited sets of variables. For example, if some correlated X and Y have a bivariate normal distribution, then regressing Y on X will give us an equation where Y is a function of X and an independent exogenous term. But so will regressing X on Y. Yet only one of these will correspond to causal structure. The same point applies to larger sets of mutually correlated variables in multivariate normal distributions.

However, there is a way to exclude these unwanted cases and so uphold a version of the explicit reduction of causation (14) proposed above. The key is that, in all the unwanted cases, the requirement of exogenous independence will be violated if we seek to *expand* the set of equations to accommodate other correlated variables.

Consider what happens in the Hesslow example if we try to expand the equations (15) to include a variable for pregnancy P (which, remember, is correlated with both B and T). If we hold fixed the probabilistic independence of B and T, then such a system of equations with exogeneous independence throughout would need to take the form:

$$B = e_B \tag{15.1}$$

$$T = e*_T \tag{15.2}$$

$$P = fB + gT + e*_P \tag{16}$$

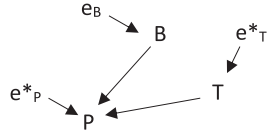

DIAGRAM 10

But the trouble now is that this augmented set of equations won't have independent exogenous terms throughout. In particular, the term e^*_T won't be independent of e^*_P, given that in truth T depends causally on P. (The term e^*_T will in effect need to compensate for how P in fact varies in ways that are independent of T.)

Let us say that a system S of equations with exogenous independence is *expandable* if, for any further variables correlated with those in S, there is a larger system of equations covering those further variables that also satisfies exogenous independence and which has S as a subsystem.

The equations (15) displaying birth control pills B and thromboses P as independent are not so expandable, since any larger set of equations embedding (15) and also covering pregnancy P will violate the requirement of exogenous independence.

Let me now turn to mutually correlated variables in multivariate normal distributions. In these cases too we can dismiss the causally misleading equations on the grounds that they are not expandable. Note here that we can usefully view the oversupply of equations with exogenous independence in multivariate normal distributions as a corollary of the way that the bridge principles sometimes fail to determine a unique causal order from a limited set of correlations. Earlier, in section 3, I pointed out that, given such underdetermination, a wider set of possible correlations would always suffice to fix a unique causal order for the variables at issue. If we now assume that such further correlations will always exist when we have mutually correlated variables in multivariate normal distributions, we can infer that only one of the original sets of equations over the originally correlated variables will be expandable, namely the one whose equational order matches the causal order fixed by the wider set of correlations. (If some other among the original set of equations were expandable, then by the "Determinism-Independence-Markov Result" (12) the so-expanded equations would yield correlations that required the bridge principles to fix a matching causal order, contrary to the hypothesis that the wider set of actual correlations plus the bridge principles do not fix that casual order.)

In line with these examples, let me now hypothesise that the requirement of expandability will always be violated by systems of recursive equations with exogenous independence that do not match causal order. If this is right, and the unwanted systems of equations can always be dismissed in this manner, then the way stands open to the following explicit reduction of causation:

(17) X causes Y if and only if it is an ancestor of Y in an *expandable* recursive structure of deterministic equations with independent exogenous terms.

On this account, a structure of causes and effects is nothing over and above a structure of variables in an expandable system of deterministic equations with independent exogenous terms. In effect, this suggested analysis of causation combines a *regularity* theory of causal *covariation* with a *statistical* account of causal *direction*. We start with a set of deterministic equations. These specify how certain variables covary deterministically in a lawlike way.[9] But this covariation is itself undirected. The covariation specified by the equations would remain the same if we reordered the equations to switch which sides the variables appeared on. The causal direction is then

added to the covariation by the requirement that the exogenous terms be probabilistically independent of each other.

11. SINGLE-CASE CAUSATION

The proposed reduction analysis of causation now promises a better hold on single-case causation. (Did Joe Blogg's smoking actually cause his cancer?)

A natural first thought is that we can now say that some particular fact Ca is an actual cause of Ea if the recursive equations relating C and E imply that, in the actual circumstances, Ea would not have occurred without Ca. However, we will then need to address the issue, familiar from David Lewis's attempt at a counterfactual analysis of actual causation, that there are cases of actual causation without counterfactual dependence, due to pre-emption, trumping, and so forth.

Fortunately, we can appeal here to the wealth of recent work which uses "causal models" to address this issue (Hitchcock 2020). These causal models posit directed relationships, standardly portrayed by arrows, between actual and possible values of variables displayed by particular situations. The existing literature then aims to formulate recipes that will allow us to read off from the models which events some given result was actually caused by or counterfactually dependent on.

The analysis of this paper complements this literature. While much progress has been made on the way causal models can help analyse actual causation and counterfactuals, it is very unclear what features of the real world these models answer to. In particular, there is no agreed account of what the arrows in the models represent (Beebee and Menzies 2020). To a large extent these relationships of asymmetric causal dependence are simply taken as given. The present paper offers a way of filling this lacuna. I would suggest that we should take there to be an arrow between two variables in a causal model just in case one is a parent of the other in an expandable system of deterministic equations with exogenous probabilistic independence.

12. "C CAUSES E"

It will be illuminating at this point to return briefly to the question of the meaning of generic causal claims of the form "C causes E" ("smoking causes cancer", "birth control pills causes thromboses", . . .). The analysis so far argues that claims of this form should be read as saying that C is an ancestor of E in a recursive system of structural equations. And I have shown how we can infer such claims, reliably if not infallibly, from correlations by using the bridge principles.

We might well wonder why we are so interested in claims of this form, given that they will standardly leave us in the dark about single-case dependencies. As observed earlier, knowing that smoking causes cancer won't decide whether Joe Blogg's cancer was actually caused by his smoking, or even whether it counterfactually depended on it. To know these things we'd need to about *all* the other equation-ancestors of cancer and about what values they had in Joe Blogg's case, and just knowing that "smoking causes cancer" is likely to leave us very much in the dark about this.

In truth, we are interested in claims like "smoking causes cancer" for a quite different reason. Even though they tell us little about single-case causal relations, they can

be a highly informative *guide to action*. It is relevant here that generic "C causes E" claims will generally come with numbers attached. They will indicate *how much more probable* C rather than not-C renders E. And this of course is just what we need to know when we are deciding whether to do C in pursuit of E.

This issue deserves far more analysis that I can give here, but at first pass we can understand quantitative generic causal claims as showing, on weighted average across the contexts consistent with the agent's information about their situation, how probable it is that doing C will realise a sufficient condition for E and not for not-E. Viewed like this, quantitative generic causal claims can thus tell agents how much reason they have to do C if they want E.

This now casts an extra light on the meaning of generic causal claims. I am currently committed to reading "C causes E" as saying that C is an ancestor of E in a system of generic casual equations. But note that not all such cases will be ones in which C on average makes a *positive* probabilistic difference to E. It might be an ancestor and yet make a negative probabilistic difference to E, or indeed make a *zero* probabilistic difference, as in faithfulness failures.

So now there is an issue about the consonance of my analysis with everyday usage. It is not obvious that we would ordinarily say that "C causes E" when C has no net influence on E, as in failures of faithfulness, and moreover it would seem positively misleading to say "C causes E" when C lowers the probability of E. Still, I do not propose to pause on these terminological points. I have identified the crucial relation of ancestry in equations with exogenous independence, and shown how this relation can be reliably though not infallibly evidenced by correlational patterns. I have then indicated how this relation might help us to understand a number of further causal relations, including actual causation, counterfactual dependence, and making a positive, zero or negative causal difference on weighted average across a type of context. Analysing how all these different causal relations map onto everyday usage is a task for another time.

13. QUANTUM MECHANICAL INDETERMINISM

The reduction of causation I have proposed so far assumes that effects are always *determined* by antecedent facts—values of dependent variables X_i are deterministic functions $F(X_1, \ldots X_{i-1}, E_i)$ of the independent variables $X_1, \ldots X_{i-1}$, and the exogenous variables E_i. At first sight this might seem inconsistent with the indeterministic nature of the world revealed by quantum mechanics.[10]

But let us not be too quick. Note that my reduction does not imply that everything is determined, only that *effects* are. Note also that it does not require that, at *every time* earlier than an effect, facts obtain that determine that effect, only that all effects be determined by facts that obtain *by the time* they occur.

This leaves it open that many of the facts that determine an effect might themselves be the outcome of chancy quantum processes. The multiple influences that contribute to the exogenous variables could still be the outcomes of chancy quantum processes, and moreover the values of those exogenous variables might only become determinate shortly before the time of the relevant effect. That would be perfectly in

line with the idea that the values of dependent variables are always deterministic functions of probabilistically independent exogenous variables.

So my deterministic analysis does leave room for indeterminism *outside* causal relations, so to speak. Still, this does not fully answer the worry. What rules out indeterminism entering into causal structures themselves? Suppose I make a bomb that is set to explode if a radioactive substance decays by a certain amount in a certain interval. If the bomb explodes, then my action will have caused the explosion. But not every event in the causal chain from my action to the explosion will have been determined by the time it occurred. In particular, the relevant radioactive decay will have been a purely chancy matter.

We can expect many causal structures to share this form. Take the stock example of smoking and lung cancer again. Perhaps the causal route from smoking to cancer proceeds via the chancy breaking of certain bonds in certain molecules. Again, this will mean that one of the steps in the causal chain running to smoking from cancer will not have been determined by the time it occurs.

At first pass, cases like these call for structural equations of a different form. In place of equations like

$$E = F(C,\ e_E) \tag{18}$$

we will need

$$\text{Chance}(E) = F\left(C,\ e_{\text{Ch}(E)}\right) \tag{19}$$

If event E is undetermined until the time when it occurs, then the variables appearing on the right-hand side of the equation (19), including all influences packed into the exogenous variable $e_{\text{Ch}(E)}$, will fail to determine a definite value for E. Instead they will fix only that E has a certain chance of occurring.

This change of equational form matters significantly to the arguments of this paper. Recall the crucial earlier Determinism-Independence-Markov Result (12) that implied that variables involved in structural equations with exogenous independence will satisfy the Markov Condition, and so in particular that correlated variables will become conditionally uncorrelated when we control for intermediaries. This result depended not just on the probabilistic independence of the exogenous variables, but also on the determinism of the equations. In consequence, recursive structures of equations some of which only fix chances rather than definite values for their dependent variables, as in (19), are no longer guaranteed by the independence of their exogenous variables to satisfy the Markov Condition.

A real-life illustration of this abstract possibility is provided by the so-called Einstein-Podolsky-Rosen ("EPR") correlations (Einstein et al. 1935). In these cases, an initial state, plus further background factors including the setting of instruments, fixes chances for various quantum measurements made on two wings of an experiment involving spatially separated particles. Quantum mechanics predicts, and experiment confirms, that these measurement outcomes will be correlated in a way that

cannot be screened off by any features of the initial parent state, even if the background factors relevant to the two measurements are probabilistically independent.

We can represent this kind of set-up by the pair of equations:

$$\text{Chance}(E_1) = F\left(C,\ e_{Ch(E1)}\right) \tag{20}$$

$$\text{Chance}(E_2) = F\left(C,\ e_{Ch(E2)}\right) \tag{21}$$

And now the relevant point is that the nondeterminism of these equations *leaves room* for the outcomes to coordinate themselves within the freedom, so to speak, left open by their nondetermination. When we have two equations *determining* two outcomes E_1 and E_2 as functions of some common cause C and two independent exogenous variables e_{E1} and e_{E2}, then the independence of the last two terms *forces* the conditional independence of E_1 and E_2 given C, as in the earlier Determinism-Independence-Markov Result. But when the only outcomes fixed by the equations are *chances* for E_1 and E_2, then there is room to evade this conditional independence. And the EPR correlations show that, once this room is made available, nature sometimes makes use of it.

Still, it is noteworthy that correlations with this non-Markov nature are effectively unknown outside the physics laboratory. It requires very carefully arranged experimental circumstances to display the characteristic features of the EPR correlations. I take it that this is due to the fact that, in the absence of such careful experimental arrangements, the parts of separated entangled systems that might display EPR-type correlations will quickly interact with different macroscopic systems that are not specifically designed to amplify the values of the entangled variables in concert. Because of this, any spatially separated macroscopic events that are influenced by different parts of entangled quantum systems will vary independently, once we hold fixed the common sources of those quantum systems.

And this then means that for practical purposes any structural equations with macroscopic dependent variables in which chancy quantum events play a role can be represented as deterministic after all. For we can now effectively rewrite equations of the form

$$\text{Chance}(E) = F\left(C,\ e_{Ch(E)}\right) \tag{19}$$

as

$$E = F\left(C,\ e_{Ch(E)},\ e_E\right) \tag{22}$$

where the final e_E is a sort of "dummy variable" representing the way in which the chance of E resolves itself into actuality. As long as we are dealing with cases, unlike the carefully arranged EPR set-up, where different chancy variables will macroscopically resolve themselves independently, we can assume that these extra chance-

realizing dummy variables in different equations will be probabilistically independent of each other.

We can usefully describe equations like (22) as "pseudodeterministic." Dependencies that appear probabilistic only because they omit the totality of determining factors are often termed "pseudo-*in*deterministic." But contrast, while equations like (22) have the appearance of determining the effect variable E, this conceals the way that the e_E terms on the right-hand side are an expression of the fact that nothing determines E until it occurs.

Still, as long as these dummy chance-realizing variables in a system of structural equations are probabilistically independent, then the equations will function just like a system of deterministic equations with exogenous independence, with the result that we can uphold the same connection between causes and correlations as before. The Causal Markov Condition with its correlative Linkage Principles will be a deductive consequence of the analysis, while the Faithfulness Condition and its correlative Unlinkage Principles can be expected to hold except in special cases of faithfulness failure.

So I now propose the following adjusted analysis of causation to accommodate the involvement of chancy events in causal structures:

> (23) X causes Y if and only if it is an ancestor of Y in an expandable recursive structure of deterministic *or pseudodeterministic* equations with independent exogenous terms.[11]

One consequence of this adjusted analysis is that the EPR relationships will not themselves qualify as causal. The equations governing the EPR outcomes are not pseudodeterministic. If we tried to put these equations into the pseudodeterministic form (22), the "dummy" exogenous variables representing the undetermined manifestation of the outcomes on the two wings would not come out as probabilistically independent.

Denying causal status to the EPR relationships seems independently reasonable. Even though the two spatially separated outcomes are connected by an unscreened-off correlation, there is good reason to deny that either causes the other. After all, the relationship between the two wings is symmetrical, and moreover there is no possibility of controlling the result on one wing by manipulating the other. As to the production of the spatially separated outcomes by their common source, there is again reason not to count this as the production of distinct joint effects by a common cause. After all, their covariation cannot be screened off by values of the source, as normally happens with joint effects of a cause. Given this, we will do better to regard the coordinated outcomes as together comprising a single effect resulting from the source, rather than two distinct effects with independent sources of variation.

If we do take this line, then the EPR correlations become counterexamples to the Linkage Principles. These Principles said that a correlation always signifies a causal link, and a conditional correlation signifies a causal link that doesn't pass through the condition. But the EPR correlations, which remain even after we condition on the

source, do not signify a direct causal link between the two outcomes, nor even an indirect causal link resulting from a common cause.

Now, as before, this violation of the Linkage Conditions is no problem for practical nonexperimental researchers. As I have observed, we can be confident that we will not meet any observable EPR correlations outside the laboratory setting. So practical researchers can continue to assume the Linkage Conditions in inferring causes from correlations.

Still, one might wonder where the EPR correlations leave my claim that my proposed analysis of causation has the Causal Markov Condition and hence the Linkage Conditions as a deductive consequence. If my analysis does indeed imply the Causal Markov Condition and therewith the Linkage Conditions, and the EPR correlations show the Linkage Conditions are not generally true, then that looks bad for my proposed analysis.

A crucial point here, however, is that the Causal Markov Condition says specifically that variables in any *causal structure* will satisfy the Markov Condition, not that all variables whatsoever will—and on my developed analysis causal structures are specifically expandable recursive structures of deterministic or pseudodeterministic equations with exogenous independence. The EPR correlations are thus not covered by this result, since, as we have seen, the equations governing the outcomes on the two wings in the EPR set-up cannot be put into the form of pseudodeterministic equations with probabilistically independent exogenous variables.

This does now mean, however, that my original Conditional and Unconditional Linkage Principles (1) and (2) were too generally formulated. As I originally formulated these principles, they specified that causal implications follow from *any* correlations between variables. We can now see that this was too ambitious. The EPR correlations show us that the relevant casual implications are only guaranteed if we are dealing with variables which are governed by deterministic or pseudo-indeterministic equations with independent exogenous variables. This qualification to the Linkage Principles might be of no practical importance, given that no EPR-type correlations ever present themselves to nonexperimental researchers, but it is needed if we want to keep the logic straight.[12]

14. THE TEMPORAL ASYMMETRY OF CAUSATION

One aim of this paper was to offer an explanation for the temporal asymmetry of causation. Given that this asymmetry has no counterpart in the fundamental dynamics of the physical world, we would like to be able to understand how it emerges.

The analysis I have developed puts me in a position to offer such an explanation. On my account, causal structures are expandable recursive systems of deterministic or pseudodeterministic equations in which every variable is descended in a certain way from a set of probabilitically exogenous variables. Now in reality, this causal ordering will line up with temporal ordering, in the sense that any variable that is causally prior to another according to my analysis will always in fact precede it in time. The variables on the right-hand sides of expandable structural equations with exogenous independence always turn out, as a matter of fact, to be temporally prior to their dependent variables.

So far this does little more than restate the temporal asymmetry of causation. I might have offered an analysis of causation in other terms, but I am still simply presenting it as a datum that causes so analysed will never succeed their effects. What we would like, however, is some further explanation of why that should be so.

The points made in the last section suggest that we might be able to appeal to quantum processes to meet this challenge. It is natural to suppose that the exogenous terms in expandable structural equations are the result of quantum superpositions resolving themselves into determinate outcomes when they interact with macroscopic systems. Given this, we can attribute their probabilistic independence to the typical unconnectedness of the processes giving rise to quantum "collapses." If we put the specially coordinated measurements of EPR-type experiments to one side, then quantum collapses will generally occur in interaction with unrelated macroscopic systems, and we can take it that any "collapses" they prompt will display probabilistic independence.

While this now offers a quantum mechanical account of the probabilistic independence of exogenous variables, it does not yet in itself take us to the temporal asymmetry of causation. We want to explain why the exogenous variables in structural equations are always *temporally prior* to the variables that depend on them. Viewing them as the outcomes of "quantum collapses" might account for their probabilistic independence from each other. But it is not immediate obvious why this should mean they must temporally precede the further variables that causally depend on them.

It is relevant, however, that apparent quantum "collapses" themselves occur asymmetrically in time. The superposition comes first, and is then followed by the collapsed state. Different interpretations of quantum mechanics of course offer different accounts of the mechanics of quantum state "collapses." Still, it is a constraint on all these accounts that they should respect the way that such manifest collapses occur asymmetrically in time, with the determinate outcomes always occurring later that the quantum superpositions that precede them. There is more to say here, but this asymmetry of apparent quantum collapses thus promises to account for the way exogenous variables in structural equations always temporally precede their dependent variables.

At the beginning of this paper I mentioned the Lewisian programme of accounting for the asymmetry of causation in terms of the "asymmetry of overdetermination." This asymmetry consists in the fact that any time will contain many independent traces of past events, but scarcely any of future events. This is a real enough phenomenon, but from my perspective it is not prior to the asymmetry of causation. As I see it, the "asymmetry of overdetermination" derives from the asymmetric nature of causation, not the other way round.

To see why, note how my account implies that the joint effects of any cause will be generally be correlated with each other. Two variables that are both correlated with a common ancestor in a system of deterministic or pseudodeterministic equations with exogenous independence will be probabilistically dependent, putting unlikely faithfulness failures aside. By contrast, nothing in my account implies that two variables with a joint descendant will generally be correlated. This means that the

joint effects of any cause will tend to occur in concert, in a way that joint causes will not. Any given cause will thus tend to be followed by a plurality of different events, each of which probabilistically indicates it. By contrast, any given effect will typically be correlated only with one identifiable precursor. This thus allows us to account for the "asymmetry of overdetermination" as an upshot of the way that causation is itself asymmetrically orientated in time.[13]

15. CONCLUSION

I have argued that the causal structure of the world arises from the way certain variables are governed by deterministic or pseudodeterministic structures of equations with exogenous independence. This analysis allows us to understand a number of features of causation, including our ability to infer causes from correlations, the grounds of actual causation and counterfactual dependence, and causation's temporally asymmetry.

Perhaps I should make it clear that that this account is intended as an a posteriori analysis of the *nature* of causation, not as any kind of conceptual analysis of the *concept* of causation. I have not sought to derive my analysis a priori from the way we intuitively think about causation, but rather have offered it as the best explanation for a range of a posteriori facts about causation, most centrally for the way causation characteristically displays itself in correlational patterns.

Some might feel inclined to object to my analysis that they can perfectly well conceive of one event causing another without the help of any exogenous variables satisfying probabilistic independence requirements. Consider, for example, a world with nothing else in it where one perfectly hard ball bumps into another and "causes" it to move (cf. Ehring 1987; Sosa and Tooley 1993, Introduction.) My response is that we might be able to conceive of such a world, but we would be conceiving a metaphysical impossibility (Papineau 1988). Given that my a posteriori analysis makes no appeal to the concept of causation, I am happy to allow that we can coherently apply that concept to imaginary situations that fail to satisfy the analysis. But the resulting description, while conceptually consistent, will be metaphysically contradictory. It will describe a set-up that violates the a posteriori nature of causation. In truth, causation depends on expandable systems of equations with exogenous independence, and will be absent from any world that lacked such complexity.

So it is no objection to my analysis that we can *conceive* of causes without independent exogenous variables. But there is a related worry. My analysis does at least require that all causal relations in the *actual* world are embedded in expandable equations with exogenous independence—and this itself might seem an overly strong and implausible claim. How can I be confident that every single effect in the world has a plurality of causes one of which is probabilistically independent of the others?

I take the points made in the last two sections to provide an answer to this query. No observable feature of the world is insulated from the impact of chancy quantum processes. This is not to deny that some prior circumstances do make others overwhelmingly likely and so come very close to determining them. In particular, we humans often go to great pains to arrange things to ensure that some specific outcome will follow. But absolute determination of any event by another at a temporal

distance is an unattainable ideal. In principle, freaky quantum events can always disrupt any outcome. Variables representing the absence of these events will thus feature among the causes of any observable outcome, and will thus provide the requisite exogenous independence, for the reasons given in the last section.

What if we conjoin all the influences that contribute to the outcome under consideration into one big determining cause? Then there will be no plurality of variables influencing the result, and so no question of whether one is probabilistically independent of the others (Hausman 1998, 215). But that is not to the point. The central thesis of this paper is not that *all* ways of grouping the causes of an effect will present it as a function of probabilistically independent factors, but rather that there is always *some* way of so grouping its causes, which will then constitute it as an effect of those factors. In my view the underlying quantum nature of reality gives us every reason to accept this thesis.[14]

NOTES

1. "Variable" can be understood as referring to a symbol on paper or in some other medium, to a function with abstract numbers as values used to model some worldly quantity, or to the worldly quantities themselves. My focus throughout this paper will be on the last-mentioned worldly quantities. Similarly, I shall understand "equation" as standing for lawlike worldly relationships between such quantities, and not for any symbolic or abstract numerical representation thereof.
2. Lewis's original explanation was given in his "Counterfactual Dependence and Time's Arrow" (1979). Elga (2000) showed that Lewis's treatment was insufficiently sensitive to thermodynamic considerations and therefore wrong to view later traces as strictly *determining* earlier events. Loewer (2007) remedied this deficiency but simply assumed without any further analysis that the "asymmetry of overdetermination" (in the sense of "the predominance of local macro signatures of the past (but not of the future)" 2007, 317) is built into the asymmetry of thermodynamics.
3. By "correlation" I shall mean any case of nonindependent probability distributions of two variables (school type/examination results, pollution/death rate, . . .) instantiated by some type of individual (schoolchildren, towns, . . .). I shall assume throughout that these are lawlike population patterns, as opposed to finite sample associations, and that they hold within some background field.
4. The more familiar coinage is directed acyclic "graph" (DAG). I have adopted "structure" instead to stress that my concern is with worldly relationships between worldly quantities, and not with the means by which we might represent these relationships. In such a causal DAS, an arrow means that the variable at the head of the arrow is a direct cause of the variable at the tail. The arrows in such a DAS are required to be acyclic in the sense that a variable can only be a causal ancestor of another if it is not also a descendant of it. Throughout this paper I shall assume that variables never reciprocally cause each other. When some coarse-grained variables seem to leave this as a possibility—for example, might not *happiness* cause *health*, and *health* also cause *happiness*?—then we should switch to time-lagged versions of these variables, as in health_{t1}, health_{t2}, health_{t3}, ...
5. The Causal Markov Condition says: In any directed acyclic structure of causal relationships, any variable will be probabilistically independent of every other variable (apart from its own causal descendants) conditional on its causal parents. (Cf. Spirtes et al. 1993, 54). A "structure of causal relationships" should be understood to include the causal relationships between any set of variables selected from reality. The Causal Markov Condition is only plausible if it is further understood to be required of such a structure that no common causes of included variables be omitted (for reasons to be elaborated in the next section). So understood, and supposing there are no further requirements on causal structures beyond these (but see footnote 12 below), the Linkage Principle (2) follows because any two causally unlinked variables can feature as parentless in a causal structure, and so must be uncorrelated, while the Conditional Linkage Principle (3) follows because, in the absence of any links between the two variables that don't involve {X}, controlling for {X} would screen off the correlation.
6. The Faithfulness Condition can be stated as: There are *no more* unconditional and conditional independencies than are required by the Causal Markov Condition. (Cf Spirtes et al 1993, 56.)

7. Thus Christopher Hitchcock's *Stanford Encyclopedia of Philosophy* article on "Causal Models" (2020) has a section ("4.4 The Identifiability of Causal Structure") about the way correlations plus bridge principles can underdetermine causal structure, but omits to mention that such indeterminacies are always in principle resolvable by further possible correlations.
8. Let us now assume that our variables, including school type S, can be measured on some quantitative scale, for example by level of school funding. I shall also simplify by assuming that all variables are measured from their means. Remember that, as explained in footnote 1, my "equations" are lawlike relationships between worldly quantities, not symbolic or abstract representations thereof, in line with my use of "variable" and "directed acyclic structure."
9. I take no view on the nature of lawlike deterministic connections in this paper. Everything I say is consistent with all the standard accounts of nomological necessity.
10. It is a moot point whether quantum mechanics is ultimately indeterministic. Collapse theories say so, but Everettianism or Bohmianism deny it. We can bypass this issue here, however, as both Everettianism and Bohmianism still need to account for the *apparent* indeterminism that occurs when quantum superpositions interact with macroscopic systems. In line with this, I shall understand quantum indeterminism as covering whatever happens in such interactions.
11. My earlier points in sections 11 and 12 about actual causation, counterfactual dependence, and the significance of generic causal claims for rational action all presupposed determinism. The admission of indeterministic causes means that these analyses need to be re-examined. That will have to be a project for another time. My hope is that the requirement of pseudo-indeterminism will mean that the earlier analyses can be smoothly extended.
12. Did I not argue earlier in footnote 5 that the Causal Markov Condition implied the original Linkage Principles (1) and (2) in full generality? But at that stage we were assuming that nothing is required of causal structures beyond including all common causes of included variables, and given this the unqualified Linkage Principles did indeed follow from the Causal Markov Condition. But now that we are restricting causal structures to expandable systems of equations with exogenous independence, the Causal Markov Condition and hence the Linkage Principles will no longer apply to EPR quantum correlations.
13. Loewer (2007) does aim to explain, in terms of the "past hypothesis," why we have "records" of the past but not the future. But his account fails to explain why we have *many* separate such records, which is what he assumes when he explains the direction of causation in terms of the asymmetry of overdetermination.
14. I would like to thank Thomas Blanchard, Julien Dutant, Toby Friend, Totte Harinen, Jenn McDonald, Davide Pigoli, Henry Taylor, and two anonymous reviewers for helpful comments on earlier versions of this paper.

REFERENCES

Beebee, H. and P., Menzies 2020. "Counterfactual Theories of Causation," in E.N. Zalta, ed., *The Stanford Encyclopedia of Philosophy*, Summer.

Blalock, H. 1964. *Causal Inferences in Nonexperimental Research*, New York: W.W. Norton.

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

——. 2002. "Against Modularity, the Causal Markov Condition, and any Link between the Two: Comments on Hausman and Woodward," *British Journal for the Philosophy of Science* 53**:** 411–53.

Ehring, D. 1987. "Papineau on Causal Asymmetry," *British Journal for the Philosophy of Science* 38: 81–87.

Einstein, A., B. Podolsky, and N. Rosen 1935. "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?" *Physical Review* 47: 777–80.

Elga, A. 2000. "Statistical Mechanics and the Asymmetry of Counterfactual Dependence," *Philosophy of Science* 68: 313–24.

Good, I.J. 1961-62. "A Causal Calculus I–II," *British Journal for the Philosophy of Science* 11: 305–18; 12: 43–51.

Glymour, C. 2004. "Making Things Happen by James Woodward," *British Journal for the Philosophy of Science* 55: 779–90.

Hausman, D. 1998. *Causal Asymmetries*, Cambridge: Cambridge University Press.

Hausman, D. and J. Woodward 1999. "Independence, Invariance and the Causal Markov Condition," *British Journal for the Philosophy of Science* 50**:** 521–83.

Hesslow, G. 1976. "Two Notes on the Probabilistic Approach to Causality," *Philosophy of Science* 43: 290–92.

Hitchcock, C. 2020. "Causal Models" in E.N. Zalta, ed., *The Stanford Encyclopedia of Philosophy*, Summer.

Hoover, K. 2003. "Nonstationary Time Series, Cointegration, and the Principle of the Common Cause," *British Journal for Philosophy of Science* 54: 527–51.

Lewis, D. 1979. "Counterfactual Dependence and Time's Arrow," *Nous* 13: 455–76.

Loewer, B. 2007. "Counterfactuals and the Second Law" in H. Price and R. Corry, eds. *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*, New York: Oxford University Press, 293–26.

Papineau, D. 1988. "Response to Ehring's 'Papineau on Causal Asymmetry'," *British Journal for the Philosophy of Science* 4: 521–25.

——. 1992. "Can We Reduce Causal Direction to Probabilities?," *Philosophy of Science Association*, 2: 238–52.

——. 2001. "Metaphysics over Methodology: Why Infidelity Provides No Grounds to Divorce Causes from Probabilities," in M.-C. Galavotti, P. Suppes, and D. Costantini, eds., *Stochastic Causality*, Stanford: CSLI Publications, 15–38.

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

Peters, J., D. Janzing, and B. Schölkopf 2017. *Elements of Causal Inference: Foundations and Learning Algorithms*, Cambridge MA.: MIT Press.

Reichenbach, H. 1956. *The Direction of Time*, Los Angeles: University of California Press.

Salmon, W. 1984. *Scientific Explanation and the Causal Structure of the World*, Princeton: Princeton University Press.

Schurz, G. and A. Gebharter 2016. "Causality as a Theoretical Concept," *Synthese* 193: 1073–103.

Simon, H. 1953. "Causal Ordering and Identifiability," in W. Hood and T. Koopmans, eds., *Studies in Econometric Method: Cowles Commission for Research in Economics*, New York: Wiley, 49–74.

Sober, E. 2001. "Venetian Sea Levels, British Bread Prices, and the Principle of the Common Cause," *British Journal of Philosophy of Science* 52: 331–46.

Sosa, E. and M. Tooley 1993. "Introduction," in E. Sosa and M. Tooley, eds., *Causation*, Oxford: Oxford University Press, 1–32.

Spirtes, P., C. Glymour, and R. Scheines 1993. *Causation, Prediction and Search*, New York: Springer.

Spohn, W. 2001. "Bayesian Nets Are All There Is to Causal Dependence," in M.-C. Galavotti, P. Suppes, and D. Costantini, eds., *Stochastic Causality*, Stanford: CSLI Publications, 157–72.

Strevens, M. 2007. "Review of Woodward, *Making Things Happen*," *Philosophy and Phenomenological Research* 74: 233–49.

——. 2008. "Comments on Woodward, *Making Things Happen*," *Philosophy and Phenomenological Research* 77: 171–92.

Suppes, P. 1970. *A Probabilistic Theory of Causality*, Amsterdam: North-Holland Publishing Company.

Woodward, J. 2003. *Making Things Happen. A Theory of Causal Explanation*, New York and Oxford: Oxford University Press.

——. 2008. "Response to Strevens," *Philosophy and Phenomenological Research* 77: 193–212.

Zhang, J. and P. Spirtes 2014. "Choice of Units and the Causal Markov Condition," in G. Guo and C. Liu, eds., *Scientific Explanation and Methodology of Science*, Singapore: World Scientific, 240–51.