The Causal Structure of Reality

David Papineau

[This is a long (double-length) version of a paper to appear in the *Monist*—and also in effect a draft of the first part of what in due course should become a book about causation. Comments are very welcome.]

[NB this is a revised version of the paper I originally posted in PhilSci-Archive in August 2021. Some mistakes in sections 13-15 have been rectified.]

1 Introduction

The aim of this paper is to develop an analysis of causation that will cast light on a number of its more puzzling aspects.

I shall initially develop this analysis in order to explain the "causal inference" techniques used by non-experimental scientists to infer causal structures from sets of correlations. The logic of these techniques has now been thoroughly codified (Spirtes et al 1993, Pearl 2000), but surprisingly none of the mainstream philosophical approaches to causation, in terms of counterfactuals, or powers, or processes, or dispositions, offer any explanation for why these techniques work.

The only philosophers who have attended to this issue are those in the minority tradition of "probabilistic theories of causation" designed to reduce causal relationships directly to correlational ones. As I shall explain, however, this approach focuses on superficial features of causation rather than its underlying nature. I shall show how the causal inference techniques are better explained not by a direct reduction to correlations by rather by an analysis of causation in terms of structures of laws with probabilistically independent exogenous terms.

A second aim of my analysis will be to explain the temporal asymmetry of causation. As has often been observed, this temporal asymmetry has no counterpart in the fundamental dynamics of the physical world. So causation must somehow emerge from the underlying dynamics, along with other macroscopic phenomena like entropy increase and the arrow of radiation.

My analysis will offer an explanation for this emergence. The requirement that structures of causal laws must have probabilistically independent exogenous terms imposes a recursive order on those structures and a consequent distinction between causes and effects. And, as a matter of observable fact, causes so analysed turn out always to precede their effects in time (Pearl 2000 sect 2.8). The temporal orientation of causation thus turns out to be due to the way that the probabilistically independent exogenous terms in causal structures always temporally precede the variables that causally depend on them. I shall suggest that this in turn is an upshot of the way that apparent quantum wave function collapses occur asymmetrically in time.

A third aim of my analysis will be to understand single-case actual causation and counterfactual dependence. Over the last couple of decades, many philosophers have appealed to "causal models" to help understand these single-case relations. However, the status of these models, and in particular of the directed "structural equations" they invoke, has remained obscure. My analysis will complement this work, by offering an independent grounding for these equations.

In his "Causation" (1973), David Lewis urged that we turn away from regularity theories of causation and seek instead to ground causation in counterfactual dependence. This paper in effect represents a reversion to the older tradition of regularity theories. A central component in my analysis is the way that causally related variables are connected by laws. However, by adding the requirement that the exogenous terms in the laws be probabilistically independent, my analysis meets the main challenge that Lewis took to face regularity theories, namely that of distinguishing, among variables connected by laws, causes from effects and symptoms. (I should say that my analysis will be neutral on the nature of laws themselves. I take it to be compatible with best-systems or other Humean accounts of laws, with Armstrongian necessitarianism, and with dispositional essentialism. I shall say nothing more about lawlikeness in this paper.)

Lewis's approach requires counterfactual dependence to be metaphysically prior to causation. In the years since his original paper, however, it has become apparent how difficult it is to articulate a semantics for counterfactuals that does not itself invoke causal considerations. This is because the evaluation of many counterfactuals seems to require us to hold fixed just those facts that are not *causally* dependent on the antecedent (Schaffer 2004). By offering an independent account of the directed "structural equations" assumed by "causal models", my analysis will open the way to an explanation of counterfactual dependence that is able to appeal to independently constituted causal relations.

Lewis himself came to recognize that his analysis of causation in terms of counterfactuals could not explain the temporal asymmetry of causation without invoking a substantial further asymmetry in time, which he called the "asymmetry of overdetermination". Others in the Lewisian tradition have refined and developed this approach to causal asymmetry.[1] 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 itself something that calls for explanation in a world with a symmetric fundamental dynamics. My analysis will have the virtue of showing how Lewis's asymmetry is itself a consequence of the asymmetric nature of causation.

The general plan of the paper is as follows.

[1] 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 states. 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)" 317) is built into the asymmetry of thermodynamics.

Sections 2-7 outline the standard procedures for inferring causes from correlational patterns.

Sections 8-11 consider the possibility of reducing causation directly to correlational patterns and explain why this faces problems.

Sections 11-15 argue that a different analysis can also explain the correlational techniques while avoiding the problems: we need to reduce causation, not directly to observed correlations, but to underlying systems of structural equations with probabilistically independent exogenous variables.

Sections 16-18 then relate this analysis to issues of single-case causation and the relevance of causation to rational action.

Finally, sections 19-21 appeal to quantum mechanical indeterminism to account for the probabilistic independence of exogenous causal terms and hence for the temporal asymmetry of causation.

2 Explaining Causal Inference

For over a hundred years epidemiologists, econometricians, educational sociologists and other non-experimental scientists have been using sophisticated statistical techniques to infer causal structures from correlational data.[2] It is an oddity, to say the least, that none of the main philosophical theories of causation casts any light on why these techniques work. What in the nature of causation allows such inferences to proceed? Why do causal structures have a distinctive correlational signature? As far as I know, nobody working on counterfactual, or regularity, or process, or dispositional theories of causation so much as asks this question.

Even those contemporary philosophers who do engage with the statistical causal inference techniques tend to avoid the issue. They are generally suspicious of explicit reductions and are happy to treat causation as a primitive relation. They then simply posit certain connections between causes and correlations as contingent truths, and turn away from any awkward queries about their metaphysical grounding.

One figure who is sensitive to the issue is Judea Pearl, the computer scientist who over the past few decades has done much to systematize the non-experimental study of causes. In response to the question of why causal structures display themselves in distinctive correlational patterns, Pearl is wont to say that this is "a gift from the gods"[3]. It is creditable

[2] This non-experimental tradition arguably goes back to Durkheim's *Suicide* (1897) and beyond. In the 1920s the geneticist-statisticians Sewall Wright (1921) and R.A. Fisher (1925) developed mathematical foundations for statistical causal inference. Their techniques were widely adopted by econometricians, including Nobel prize winners Trygve Haavelmo, Jan Tinbergen, Ragnar Frisch and H.A. Simon (see Pollock 2014), and also by social scientists, including Paul Lazarfeld (Lazarfeld and Rosenberg 1955) and H.M. Blalock (1967). More recently the influence of computer science has led to further codification and widespread applications: see Spirtes et al 1993, Pearl 2000, Peters et al 2017.

[3] Pearl 2017 9, 2018 116.

that Pearl recognises the puzzle, and understandable that as a non-philosopher he is happy to thank providence for an observable signature of causal structures. But his response highlights the manifest challenge to metaphysicians. Can it just be a contingent coincidence that the casual structures line up so nicely with the correlational patterns?

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 (1961-2) and Patrick Suppes (1970) all offered variations on this theme, and more recently Wolfgang Spohn (2001), Clark Glymour (2004), and Gerhard Schurz and Alexander Gebharter (2016), and myself (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 reduction proceeds too quickly. Because it ties causation directly to correlations, 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 analysis. The idea goes back to H. A. Simon and others in the 1950s and 60s (Simon 1953, Blalock 1967). It seeks to reduce causation to underlying structural equations with probabilistically independent exogenous terms rather than directly to surface correlations. By appealing to this underlying structure, this analysis will prove better able to accommodate both faithfulness failures and single-case causation. While the importance of probabilistically independent exogenous terms is often enough mentioned by practising non-experimental scientists, it has been largely ignored by philosophers (though see Cartwright 1989, Papineau 1991, Hausman 1998). I shall show that this independence holds the key to the success of the correlational inferential techniques.

The probabilistic independence of exogenous terms in causal structures will also turn out to account for the temporal asymmetry of causation. At the end of this paper I shall show how the temporal orientation of causation is due to the way that the probabilistically independent exogenous terms in causal structures always temporally precede the variables that causally depend on them.

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 options described above, namely, viewing connections between causes and correlations non-reductively as fortunate contingencies, or alternatively reducing causes directly to correlations. 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, this does forge some link between correlations and causes, and thereby casts some light on the causal inference techniques (Hausman and Woodward 1999). But at the same time, it shares deficiencies with both the non-reductive and reductive options mentioned above. On the one hand, it is a moot point whether it can cope with faithfulness failures any better than the reductive option, given its equation of causation with correlation under intervention (Strevens 2007 2008, Woodward 2008). And, on the other, it does not give a full explanation for the distinctive correlational signature of causes, since it appeals circularly to causal conditions in defining an "intervention". 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 A Simple Example

Let me start by illustrating the kind of statistical techniques at issue with a simple example. Suppose educational sociologists studying the effects of high schools on examination results discover that there is a positive correlation between the type of school attended (S) and examination results (E). The children attending well-funded schools tend to score better in school-leaving examinations than those from less affluent ones.

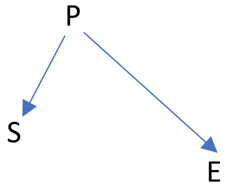
At first pass, this is some evidence that better school funding causes higher examination results.

But now suppose that the correlation disappears when we "control" for parental income. Among children with the same level of parental income (P), the children in poorly-funded schools do as well as those in highly-funded schools. As it is often phrased, P "screens off" E from S.

This further evidence now argues that school funding does not cause higher examination results after all, but that their initial correlation was rather due to their both having a common cause, higher parental income. The children in well-funded schools do better, not because of the schools, but because of other advantages deriving from rich parents. The association of schools with examination results was simply due to well-funded schools containing more such already-advantaged children.[4]

The overall evidence thus suggests the following casual *directed acyclic structure* ("DAS" henceforth):

(1)



[4] I make no claims for the accuracy of this example. As it happens, the sociological evidence in the USA and the UK does suggest that school funding has surprisingly little effect on academic performance, though the issue remains much debated. See for example Dearden et al 2002 Wenglinsky 2007.

In such a causal DAS, an arrow means that the variable at the head of the arrow causes the variable at the tail.[5] The arrows in such a DAS are required to be acyclic in the sense that that a variable can only be a causal ancestor of another if it is not also a descendant of it—where "ancestor" and "descendant" have the obvious definitions in terms of arrows of causal influence.[6]

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. Such claims do not wear the meaning on their sleeve, nor is it obvious how to make them precise. 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.

This educational example of an inference from correlations to causal structure illustrates the philosophical challenge I wish initially to address. What is it about the causal relation that allows such inferences to proceed? As I said, none of the major metaphysical theories of causation so much as raise this issue.

4 Correlations

Since they will figure centrally in what follows, it will be worth saying something more about *correlations* at this point. By their nature, correlations generalise over a certain type of spatio-temporal particular. For example, the particulars might be *schoolchildren*, or *towns*, or *smokers*, or *businesses*, or *ecosystems*, or any kind of repeatable such item. We are then interested in probabilistic patterns of covariation between the values of certain variables possessed by those entities. Do a child's *examination results, school funding, and parental income* predict each other? Do a town's *wealth, literacy, and number of doctors* predict each other? Do a cigarette smoker's *number a day, level of air pollution, and lung condition* predict each other? And so on.

When I say two variables X and Y are correlated, I simply mean that their probability distributions are not independent. Their joint probability distribution Pr(X, Y) is not the product of their separate probability distributions Pr(X) and Pr(Y)). This means that the probabilities of some values of Y are sensitive to some values of X. Knowing the value of X is of some predictive significance for Y. Note that this requirement is symmetric. If X is informative about Y, then Y is informative about X.

[5] 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. Relatedly, "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.

[6] 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$_{t3}$, . . .

For two dichotomous variables A and B, correlation is simply the requirement that Pr(A&B) ≠ Pr(A)Pr(B). A and B occur together more or less often than you'd expect given their separate probabilities of occurrence. For linearly related real-valued variables, correlation is equivalent to a non-zero Pearson correlation coefficient. But correlations as I shall understand them are not restricted to just these cases. We can have non-independent probability distributions for variables with any ranges of values displaying any patterns of dependence.

When speaking of correlations in this paper, I shall always mean *population* correlations, underlying lawlike tendencies for certain types of result to occur together in a certain type of situation. Population correlations in this sense are to be distinguished from *sample* correlations. The latter are simply a finite count of how often different values of different variables occur together in some finite sample of children, towns, or whatever. Such a sample correlation can well diverge from the underlying population correlation, due to the vagaries of finite sampling.

Of course we have no epistemological route to population correlations except via sample correlations. The business of inferring population statistics from sample statistics is the subject of statistical inference. I shall say nothing about statistical inference in this paper.

We need to think of correlations as holding within a *background field*. For example, a correlation between school type and examination results won't hold for all children, whatever their circumstances, but children of a certain type, fixed by what we are taking for granted. Thus it might be taken as given that a system of social benefits is in place, that all children have access to a television, that all teachers have a tertiary educational qualification, that examination results are not determined by bribery, and so on . . . And in general any correlational study will assume that background circumstances have been fixed in ways that ensure the stability of the patterns being investigated.[7] (Later I shall be more specific about exactly which stable patterns matter for causal structure. It will turn out that underlying equations and probabilistic independencies are crucial, but that certain further features of correlations are not.)

5 Bridge Principles

As the example of school funding and examination results illustrated, it is natural to draw causal conclusions from correlational data, and much work in the non-experimental sciences does exactly that. Still, how exactly is the trick done? As we are often reminded, correlation is not causation. For a start, correlational relationships are symmetrical, while causal relationships are not. So what assumptions might allow researchers to move from the former to latter?

[7] In real correlational studies, this kind of specification will standardly be left implicit. While researchers might take care to ensure that their samples are representative of some group—Californians, say—they won't normally pause to specify which properties of that group matter, and will at best identify them implicitly—as those properties required for their findings to hold good. (Often enough, though, this issue is forced into the open by questions of "external validity"—how far should we expect the findings for California to apply elsewhere, and if not why not?)

Recent work in the "Bayesian network" tradition has done much to codify the assumptions that enable non-experimental scientists to extract asymmetrical causal structures from sufficiently rich set of correlations (Spirtes et al 1993, Pearl 2000, Peters et al 2017). In this section and the next two, I shall articulate these assumptions 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 in 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.

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

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

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

(3) 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 that by-pass those controlling variables.[8]

[8] The Causal Markov Condition says:

(4) 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 here be understood to include any set of causal relationships abstracted from reality. The Causal Markov Condition is only plausible if such structures are further understood to require that no common causes of included variables be omitted (for reasons elaborated in section 7 below). So understood, and supposing there are no further requirements on causal structures beyond these (but see section 19 below), (2) follows because any two causally unlinked variables can feature as parentless in a causal structure, and so must be uncorrelated, while (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.

These two principles are already enough to facilitate certain inferences from correlations to causes. Consider our educational example again. P, S and E were all pairwise correlated. So, by the Linkage Condition (2), they must all be pairwise causally linked.

But the principles (2) and (3) only take us so far. They tell us we can infer causal linkages from correlations. However we also want to be able to infer *absence* of causal linkage from *absence* of correlations. In the literature this is normally accommodated via a "Faithfulness Condition", but once more we will do better to focus on two perspicuous consequences.

First, an *Unlinkage Principle*:

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

Second, a *Conditional Unlinkage Principle*:

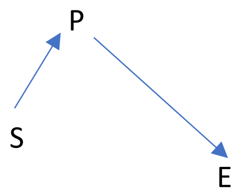
(6) If two correlated variables become conditionally uncorrelated when we control for other variables {X}, then they are not causally linked by any chains of variables that do not contain any of {X}.

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 by-pass those controlling variables.[9]

In our educational example, P screens off E from S. So, by the Conditional Unlinkage Condition (6), the link between S and E must go via P.

This now further narrows down the causal possibilities. True, this does not yet uniquely determine the structure (1), with P as the common cause of S and E, that I initially suggested as the natural causal interpretation of the correlations. For, even given the four posited principles, the given correlations are also consistent with these two further DASs:

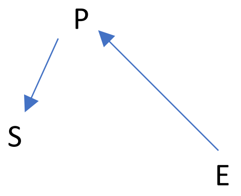
(8)



[9] The Faithfulness Condition can be stated:
(7) There are *no more* unconditional and conditional independencies than are required by the Causal Markov Condition. (Cf Spirtes et al 1993 56.)
(This principle is so-called because it requires probabilistic independencies to be *faithful* to the underlying causal structure. If an unconditional or conditional correlation is zero, then that must be because there is no corresponding causal link.)

and

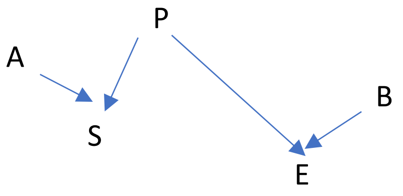
(9)



Still, these three structures are the only options consistent with the correlations according to the bridge principles

What about the unresolved choice between (1), (8) and (9)? This can’t be decided by the correlations between P, S and E, but it might well be resolved if we knew the correlations between these and some further observed variables. Suppose for example that we observed some A that was correlated with S but independent of P, and also some B that was correlated with E but again independent of P. Then we would be led to conclude that the structure of equations must be:

(10)



The logic here would be that A and P must be unlinked causes of S, since they aren’t correlated with each other but are both correlated with S, and similarly that B and P must be unlinked causes of E.

6 The Power of the Bridge Principles

I shall refer to (2), (3), (5) and (6)—equivalently the Causal Markov and Faithfulness Conditions—as the “bridge principles” henceforth. (In summary form, to repeat, these simply say that two variables are causally linked if and only if they are correlated, with the correlations being screened off if and only if we control for causally linking intermediaries.)

We have just seen one example in which these bridge principles suffice to determine a causal order among a set of correlated variables. While the correlations among our initial three P-S-E variables left their causal relationships underdetermined, the indeterminacy was resolved when we brought in correlations with two further variables.

This example illustrates a principle that can be proved in full generality. Whenever the correlations between some set of variables do not suffice for the bridge principles to fix their causal relationships uniquely, there will always be possible correlations involving further possible variables that will so suffice.[10]

[10] Theorem 4.6 Spirtes et al 1993 94.

Let me observe at this point that when the bridge principles do so fix causal order, they do so without resorting to any information about the temporal ordering of the relevant variables. Yet we can expect that, when they do 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[11]. 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.)

7 Including Common Causes

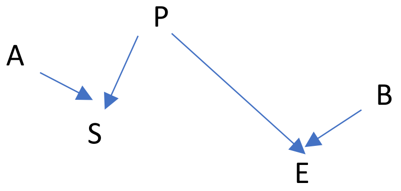
So the bridge principles allow us to infer unique causal structures from sufficiently rich sets of conditional and unconditional correlations. This might suggest the reductive idea, associated with "probabilistic theories of causation", that there is nothing more to causal relations than the patterns 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.

[11] Thus Christopher Hitchcock's *Stanford Encyclopedia of Philosophy* article on "Causal Models" (2018) 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. Even more strikingly, James Woodward's *Stanford Encyclopedia of Philosophy* article on "Causation and Manipulability" (2016) has a section ("5. Structural Equations, Directed Graphs and Manipulationist Theories of Causation") in which he offers a purported example of two different causal structures that "imply the same exactly the same facts about the patterns of correlations that obtain among the measured variables", when in truth one implies a conditional null correlation that the other doesn't.

I shall explore this idea further in the next section. But first an immediate issue must be addressed. Our examples so far, and the general theorem mentioned in the last section, have all involved the use of the bridge principles to infer causal conclusions from correlations among some *limited set of variables* (from some coarse-grained "model" of reality, as it is often phrased). 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.

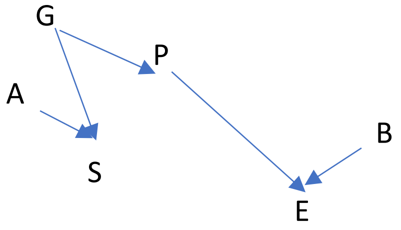
This worry is by no means an idle one. Suppose that, in the way described earlier, the correlations among some {A, B, S, P, E} determine:

(10)



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 structure would be:

(11)



However, while this kind of reversal is certainly possible, its significance is 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 probabilistic 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.)

These theoretical points prompt some obvious practical questions. Even if it is taken as given that the underlying reality satisfies the bridge principles[12], will non-experimental researchers ever know that they have included enough common causes in their analysis? This is of course a real worry. Practical researchers seeking to derive causal conclusions from survey-based correlational premises are always open to the worry that they have failed to include all “confounding variables” in their analysis. Still, it is not clear that this worry cannot be assuaged by systematic enough research. We will do well to remember the real-life history of the smoking-cancer link, in which researchers painstakingly showed that the correlation remained even after controlling for all plausible candidates for common causes, and were in this way able to mount a convincing case that the original correlation was genuinely causal. This case argues that it will often enough be practically possible, given thorough research, for non-experimental researchers to identify the true causal relationships between variables.[13]

8 Not A Gift from the Gods

I take myself now to have outlined the general logic that non-experimental researchers use to infer causal structures from correlational premises. When they draw causal conclusions from correlational premises, they do so by applying the bridge principles to unconditional

[12] The idea of “underlying reality” satisfying the bridge principles needs to be understood in terms of a level of variable inclusion at which the bridge principles are satisfied and beyond which they stay satisfied. This understanding in effect equates real causal structure with causal structure *in the limit*. Note that some such limiting notion of causal structure will be needed anyway if causation is dense, and direct causation at a coarse-grained level therefore always becomes indirect at finer levels.

[13] An alternative way for researchers to deal with the danger of confounding variables is of course to conduct a *randomized trial*. Instead of carefully surveying all possible common causes of putative cause X and effect Y, they forcibly decorrelate X from other causes of Y by experimentally assigning it to subjects at random, with the aim of ensuring that any remaining cause-effect correlation will be a genuine causal one. For more on the logic of randomized trials see footnote 18 below.

and conditional correlations among sets of variables, on the assumption that they have made these sets inclusive enough not to omit any common causes of those same variables.

As I said, there is room to question whether the bridge principles are unexceptionally true. I shall consider some such queries shortly. Still, unless we are willing to dismiss a vast body of well-respected research as groundless, we need to accept that the bridge principles contain at least some approximation to reality.

So let us now return to the question raised earlier. What in the nature of causation accounts for our abililty to infer causal structures from correlational structures in the way licensed by the bridge principles? In short, why does causation display itself in correlational signatures?

One possibility here would be to view the bridge principles simply as contingent truths. One this view, causal facts are one thing, correlational facts another. In the actual world we discover that the two sets of facts line up together, but there is no metaphysical reason why this should be so. As far as their natures go, causal and correlational facts are not guaranteed to march in step. It would be metaphysically possible to have the causal patterns without the correlational ones and vice versa.

This is the attitude expressed by Judea Pearl 's offer of thanks for "a gift from the gods". We should count ourselves lucky that we live in a universe where the causal arrows happen to have a correlational signature. The gods didn't have to arrange things like that. They could equally have allowed the causal and correlational patterns to come apart.

I find this picture difficult to take seriously. It would be like saying that temperature and molecular mean kinetic energy are two different physical quantities that just happen to go together. As far as their natures go, they could well have come apart. It just so happens that in this world they always have the same value.

I find it no more plausible that the matching of causal and correlational patterns should be a contingent coincidence than that the matching of temperature and mean kinetic energy should be so. It would beggar belief that the nature of causation should be one thing, and the correlational signature of causation quite another, with their coincidence admitting of no further explanation.

9 A Neo-Probabilistic Theory of Causation

If we are to avoid positing a brute coincidence, we need some metaphysical analysis of causation, some account of its nature that might explain why it displays itself correlationally.

The most obvious move at this point would be to hold that directed causation simply *is* correlational structure—to *reduce* causal structure directly to correlational structure. We have seen how sufficiently rich structures of correlationpromise to deliver the causal facts. So perhaps the causal facts are nothing over and above such correlational structures. On this view, the matching of causal and correlational patterns would be no gift from the gods. The two sets of patterns march in step because they are in reality a single structure described in two different ways. Not even the gods could have arranged for them to come apart. The bridge principles would not be contingent, but metaphysically necessary. They would be simply fall out of the way causal structures are constituted by correlational ones.

The original probabilistic reductions of causation, proposed in the middle of the twentieth century by Reichenbach (1956), Good (1961-2) and Suppes (1970), were all versions on this theme:

(12) 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 by-pass 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:

(13) 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 (13) follows from the bridge principles. 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.

10 The Bridge Principles Examined

Attractive as this neo-probabilistic reduction of causation might appear at first sight, it is flawed as a metaphysical analysis of causation. It locks onto the 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. Take the connection between smoking and lung cancer. I take it that the bridge principles in conjunction with actual empirical correlations have satisfactorily established 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 empirical correlations courtesy of the bridge principles, 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 neo-probabilistic reduction. Generic claims like *smoking causes lung cancer* only give us partial information about the causal links between smoking and cancer. A full analysis of causation will need to show us how to uncover this extra structure. I shall be proposing an account of this extra structure in what follows.

Associated with this deficiency are queries about how far the bridge principles are generally true. The literature contains challenges to all the bridge principles. Some of these are relatively superficial, and can be satisfactorily parried by the neo-probabilistic theory of causation just outlined. But objections to the Faithfulness Condition cannot be so easily dealt with. I shall first quickly deal with challenges to the other bridge principles in this section, before turning to the Faithfulness Condition in the next.

Let me start with the two conditions that allow us to infer causal links from correlations. First is the simple Linkage Principle (2)—correlation implies causal linkage. The standard objection is that plenty of everyday correlations seem to owe nothing to causal linkages. The annual averages for bread prices in London and water levels in Venice have been correlated ever since records began, yet this is no reason to suppose that one causes the other or that they have a common cause.

This kind of case has been widely discussed (for example by Sober 2001, Hoover 2003, Zhang and Spirtes 2014) so I shall deal with it briefly. A standard response is that correlations like these can be put to one side due to their non-standard construction. As I explained earlier, correlations signify covariation among the properties within a certain kind of particular instance (children, towns, . . .) When we seek to infer causation from such correlations, we take it that there is not also not systematic covariation among properties *across* instances. For example, in the earlier analysis of examination results we implicitly assumed that given children's parental incomes did not systematically co-vary with other children's parental incomes. (If we wanted to probe the causal significance of such cross-child covariation, we would need different units: groups of children, rather than single ones.) The bread-prices-water-level example violates this requirement of no cross-instance covariation. The instances are years, and the bread price in one year co-varies with that in the previous year, and similarly with the water levels, thus giving us a bread-water correlation for particular years that derives entirely from the monotonic increases in the two separate time series. Given this, it is open to defenders of the Linkage Condition to specify that it applies only to correlations that do not arise solely because the values of properties involved are systematically connected across instances.

I now turn to the Conditional Linkage Principle (3), which requires unconditional correlations to disappear when we control for intermediary causal links. Standard counter-examples cite common causes that supposedly do not screen off correlations among their joint effects. For example, Wesley Salmon (1984) argues that when a moving billiard ball hits a stationary one, the subsequent trajectories of the two balls are tightly correlated, yet this correlation is not screened off by the common cause, their impact. Nancy Cartwright (2002) similarly offers an example of a chemical factory where a pollutant is correlated with the production of a chemical but nothing screens off the correlation.

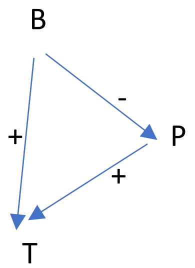
Again, cases like this have been widely discussed (Hausman and Woodward 1999, Hofer-Szabó et al 2013, Schurz 2017). The normal response is that the lack of screening-off in such examples is due to the common cause being under-described. If the precise angle of the impact were given, or the details of the factory's operation, then this would render the joint effects irrelevant to each other. Proposed counter-examples like Salmon's and Cartwright's can thus be dealt with as violating the requirement that we are dealing with variables that do not omit any common causes.

A different kind of counter-example to the Linkage and Conditional Linkage Principles involves non-local quantum correlations like those between measurements on spacelike separated entangled particles ("EPR" correlations henceforth, after Einstein, Podolsky and Rosen 1935). As normally understood, these correlations do not signify any casual links: the measurements do not cause each other, nor are they due to any common cause. I shall put such quantum correlations to one side for the moment. In section 21 below I shall say more about their relation to the Linkage and Conditional Linkage Principles.

11 The Failure of Faithfulness

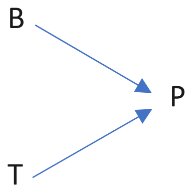
Now for counter-examples to the Faithfulness Condition. The problem cases here involve one variable causing another via two different paths, with the positive influence on one path cancelling out the negative influence on the other, resulting in a null correlation between the cause and effect. The classic example, due to Hesslow (1976), supposes that the direct positive influence of birth control pills (B) on thromboses (T) is precisely cancelled out by its negative influence through blocking pregnancies (P) which themselves conduce to thromboses, as in the causal DAS (14) below. The overall result would then be that thromboses are no more common among women who take birth control pills than among those who don't. (This would be a counterexample to the simple Unlnkage 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.)

(14)



Such "failures of faithfulness"[14] present a direct challenge to the neo-probabilistic account of causation. In Hesslow's example, birth control pills B and thromboses T are overall unconditionally uncorrelated, but both are correlated with pregnancy P—which according to the bridge principles unequivocally implies this fallacious causal structure instead of the real set-up:

(15)



Now, it is true that such perfect cancelling out would always be an unlucky freak. And this perhaps argues that we can dismiss the possibility when we are engaged with the practical business of inferring causes from correlations in real life.[15] But this dismissal is not acceptable if we are aiming at a metaphysical reduction of causation of the kind essayed by the neo-probabilistic account of causation. For this account says that causal structure is *nothing but* correlational structure, and so it needs to hold that cases where they come apart are not just unlikely, but downright *metaphysically impossible*. And the trouble is that cases like Hesslow's do not seem at all metaphysically impossible, however unlikely they may be.

12 Structural Equations

In order to deal with failures of faithfulness, we need to turn to structures that lie somewhat deeper than the correlations we have focused on so far, namely the *structural*

[14] A simpler example of faithfulness failure is due to Pearl (2000 48). A bell rings if two fair coins show the same face. The outcome of each coin toss is a cause of the bell ring, but is uncorrelated with it. This example too hinges on a sort of happenstantial cancelling-out—the equation determining the bell ring is precisely structured so the influence of one coin masks the influence of the other—and accordingly calls for the same kind of treatment as Hesslow's example.

[15] A nearby danger, however, is a real issue for empirical researchers. Even if exact cancelling of population correlations would be a freaky coincidence, *approximate* cancelling is all too likely to mislead researchers who have no alternative but to estimate populations independencies from sample statistics.

equations assumed by such traditional statistical methods as analysis of variance, regression analysis, and combinations thereof. Let me focus on the familiar methods of linear regression analysis. Go back to our original study of schools and examination results. The traditional way for educational sociologists to deal with this would be to posit these equations:

(16.1) $P = e_P$

(16.2) $S = aP + e_S$

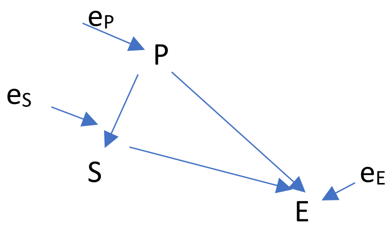
(16.3) $E = bP + cS + e_E$

(To repeat: P = parental income; S = school funding; E =examination results)[16]

The equations represent deterministic relationships. The subscripted rightmost e-terms are called "exogenous variables" and typically represent influences beyond those involved in the observed correlations. I shall call the other terms on the right-hand side of equations the "independent" variables and the terms on the left "dependent".

The above equations are *recursive* in the sense that they can be placed in an order such that no term appears as an independent variable unless it has appeared as a dependent variable in a previous equation. This means that the structure of the equations can be rendered by a directed acyclic structure as follows.

(17)



Note that this is a different kind of DAS to those introduced earlier. Where the earlier DASs were constituted by *causal* relations, this new one simply involves a recursive ordering of *equations*. Let me thus distinguish between "*equation-DASs"* and "*cause-DASs*". Much of what follows will involve the relation between these two kinds of DASs.

The coefficients a, b, c attaching to the independent variables in the regression equations (16) measure the extent to which the dependent variables vary in response to changes in the independent variables. They capture how much, if at all, the dependent variable

[16] 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 throughout that all variables are measured from their means. I shall not further standardize, however, to give all variables unit variance, as in "path analysis", as this would obscure the way in which the non-standardized slopes, and more generally the forms, of structural equations are typically more robust that the variances of their variables, as explained in section 16 below. I should also make clear that, in line with my use of "variable" and "structure", my "equations" are lawlike relationships between worldly quantities, not symbolic or abstract representations thereof.

"wiggles" when a given independent variable "wiggles" and the other independent variables are held constant.

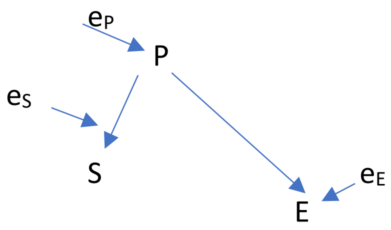
In our example, we are supposing that examination results E don't co-vary at all with schooling S once parental income P is held constant. So then the regression coefficient c will be zero, and the equations will have the simpler structure:

(18.1) $P = e_P$

(18.2) $S = aP + e_S$

(18.3) $E = bP + e_E$

(19)



Now, given the recursive structures of equation sets like (16) and (18), it is of course very natural to give them a *causal* interpretation. Those who work with regression equations typically read them as implying that the variables at the tails of the arrows are direct causes of those at the heads, and that variables not so connected by arrows are not so directly casually linked.

Still, it is not clear that anything said so far *justifies* this kind of causal reading. After all, if the equations in (16) and (18) are really *equations*, what is to stop us transforming them so as to give them a different recursive ordering? For example, viewed purely as a set of equations, (18) could happily be rewritten as:

(20.1) $S = e^*_S$

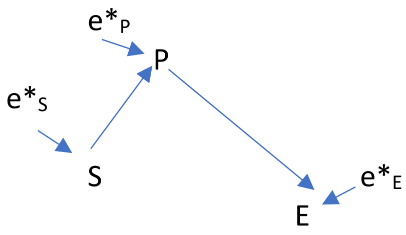
(20.2) $P = 1/aS + e^*_P$

(20.3) $E = bP + e_E$

(with $e^*_S = ae_P + e_S$, $e^*_P = -e_S/a$).

This would then give us the following alternative equation-DAS:

(21)



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, with S as having no direct causal influence on E except via P.

So—what tells us to do things the first way rather than the second? On the face of things, the equations themselves, taken purely as equations, would seem to leave both options open.

Now of course in our particular example we have independent reason to reject the second way of understanding the causal structure. After all, as observed earlier, both common sense and temporal ordering tell us that, if there is a causal influence, it will go from parental income (P) to schooling (S), rather than vice versa.

Even so, there is no need to resort to such prior knowledge to dismiss the second version of the equations. The approach to causation embodied in regression analysis has another way to select the first structure of equations as giving the right causal picture, even without any prior information about possible causal ordering.

13 Independent Exogenous Terms

The key idea is that the ordering of variables in a set of equations will capture causal structure *only if the exogenous variables are probabilistically independent*. This idea was commonplace among econometricians and sociometricians in the middle of the last century.

Note how the requirement of exogenous independence promises to decide between the alternative causal hypotheses offered by equations (18) and (20). If P causes both S and E, then the exogenous variables 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 then provides a rationale for taking the first set of equations rather than the second to represent casual structure. If in truth the exogenous variables e_S, e_P, e_E are probabilistically independent, then this argues that the first version has the causal structure right. And by the same coin, the second must then get the causal structure wrong, since the exogenous variable e^*_S in (20.1) is a linear function of the exogenous variables e^*_P in (20.2) plus another term independent of e^*_P, and so cannot itself be probabilistically independent of e^*_P.

In this section, I shall develop the idea that the probabilistic independence of the exogenous terms is a *necessary* condition for a recursive system of equations to capture causal structure. In the following section I shall then show how this condition accounts for the ability of the bridge powers to identify causal relationships. After that I shall ask whether the probabilistic independence of the exogenous terms in a recursive system of equations is a *sufficient condition* for the equation to capture causal structure. I shall show that a qualified version of this claim can be defended, and that this this opens the way to a fully reductive analysis of *X causes Y*.

For the moment I shall assume that all the variables within any causal structure are connected by deterministic equations. Quantum mechanics gives us reason to doubt that this is true. In section 19 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.

By way of initial support for the claim that exogenous independence is a necessary condition for a system of equations to capture causal structure, note how this requirement is built into the use of structural equations as a tool for prediction and explanation. Thus consider once more the equations

(18.1) $P = e_P$

(18.2) $S = aP + e_S$

which present school type S as a function of parental income P. Given some value for P, we would naturally use this to infer that expected school funding is aP. 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 the value of P in specific cases 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 (18.1-18.2), the rearranged

(20.1) $S = e^*_S$

(20.2) $P = 1/aS + e^*_P$

If e_S was independent of P in the original (18.1-18.2), then e^*_P won't be independent of S in (20.1-20.2) – remember that $e^*_P = -e_S/a$ – and so now we can't infer that 1/aS will be the average value of P given some value of S. The way P varies around 1/aS 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 1/aS.)

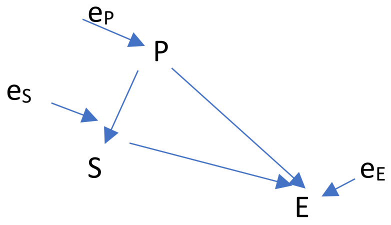
This illustrates how the probabilistic independence of exogenous variables in systems of structural equations ensures that the dependent variables are due to influences that can be factorised into independent sources. It is natural to view this as manifesting the way that the equations capture causal structure. The causal independence of the exogenous terms is displayed by their probabilistic independence, and the status of the other variables as effects is manifested by the way they are functions of these factorizable influences.

So far I have illustrated the idea in a maximally simple case with just two non-exogenous variables. But the idea that causal structure displays itself in exogenous independence can be applied in 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 (16):

(16.1) $P = e_P$

(16.2) $S = aP + e_S$

(16.3) $E = bP + cS + e_E$



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 causally dependent variable will have an associated exogenous variable that is independent of the other variables it causally depends on. This reflects the assumption, presupposed by the explicit probabilistic reduction of causation (13) I derived earlier 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 single-case causation.

I have been using linear regression analysis to illustrate the idea that causal structure implies probabilistically 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

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

Then in general we can take it to be a condition on these equations capturing causal structure that the exogenous terms be probabilistically independent. Just as with the linear regression examples I have used, the independence of the exogenous terms is naturally viewed as reflecting an underlying causal structure. Each dependent variable X has its values fixed by its independent variables and its own exogenous term. The latter adds some X-specific variation to the value for X fixed by the independent variables, and so displays X as an effect of the terms on the right-hand side of its equation.

So I now propose the following requirement on causal relationships:

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

14 Recovering the Bridge Principles

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 set of dependent variables $X_1, \ldots X_n$, 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:

(24) If the exogenous terms $E_1, \ldots E_n$, are all probabilistically independent, then any variable will be probabilistically independent of every other variable (apart from its descendants) conditional on its parents (where "parent" and "descendant" signify the obvious relations 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 round, 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 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 are independent.

Still, when we combine this theorem with the requirement (23) that causation implies recursive equations with exogenous independence, then the theorem does imply the *Causal* Markov Condition— every variable in a causal structure will be independent of every non-descendant given its parents. And this Causal Markov Condition implies that causal structures will satisfy the Linkage and Conditional Linkage Principles that play so central a role in accounting for the ability of empirical researchers to draw causal conclusions from correlational premises: correlated variables in a causal structure must be causally linked—the Linkage Principle—and 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—the Conditional Linkage Principle.

So the condition on causation (23) 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 it does not simply *posit* that causal structures will satisfy these conditions. Rather it *derives* this from the proposed connection between causation and systems of recursive equations with independent exogenous variables.

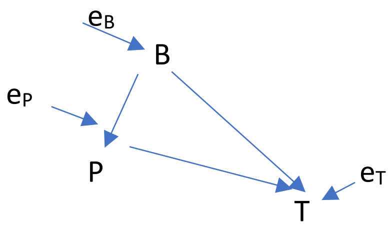
It is also noteworthy that the Faithfulness Condition does *not* follow from condition (24). That is just as it should be. As we saw earlier, it is highly implausible to suppose that the. Faithfulness Condition is built into the metaphysical nature of causation, given that the kinds of cancelling-out causal structures that give rise to them seem metaphysically perfectly possible. 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 and resulting equation-DAS.

(25.1) $B = e_B$

(25.2) $P = aB + e_P$

(25.3) $T = bB + cP + e_T$



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 imply 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.

At this point, it will worth saying something about backgrounds fields. In section 4 above I observed that projectible population correlations will be relative to a background field. The same applies to equation-DASs. (From now on it will be convenient to read "equation-DASs" as implying independent exogenous variables, in recognition of the fact that it is a condition on their conventional ordering that the ordering should respect exogenous independence.) The equations and probabilistic independencies that constitute any such equation-DAS will only hold good as long as certain background conditions are held fixed. For example, as before, we can expect the determination of examination results to work differently once we move away from contexts with a system of social benefits, access to televisions, teachers with tertiary educational qualifications, no bribery, and so on.

It is worth noting, however, that the background fields for equation-DASs will generally be less demanding than those for any given set of correlations between their variables. This is because the precise values of the correlations between variables in an equation-DAS depends not just on the equations and the exogenous independencies, but also on the amount of *variation* in the exogenous terms. For example, the extent to which school type is coupled to parental income in our example will depend not just on the linear dependencies

(16.1) $P = e_P$

(16.2) $S = aP + e_S$

but also on the extent to which e_P and e_S vary in the population under study. The greater the variation in parental income P by comparison with the other influences on S, the more S will be correlationally tied to P.[17] The linear dependency (16.2) itself, however, can be expected to be more robust than the extent of such variation. There is no preordained

[17] Under the standard assumptions of linear regression analysis, the correlation $r_{P,S}$ between P and S is related to the regression coefficient a by

(28) $r_{P,S} = a \times \sqrt{var(P)}/\sqrt{var(S)}$

It is worth noting here that failures of faithfulness, as opposed to the magnitude of specific correlations, depend only on the equation-DASs themselves, and not on the amount of exogenous variation. In equations of form (26), for example, faithfulness failure is guaranteed by $a + bc = 0$, whatever the variances.

reason why more or less homogeneity with respect to parental income, social benefits, television access, and so on, should affect the functional relationship between school type and its determiners. And so we can expect equation-DASs to hold good across wider sets of circumstances than the specific observed correlations that manifest them. The equations and exogenous independencies involved in a given equation-DAS will not automatically break down simply because we shift to a context where the variances of the exogenous terms alter.[18]

15 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:

(26) X causes Y *if* and only if X 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 yield 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 (26) representing the real causal structure, we will also have

(27.1) $B = e_B$

(27.2) $T = e^*_T$

[18] Moreover we can expect the different equations in an equation-DAS each to have their own background fields, each less demanding than the conjunction of those fields required for the whole DAS. This point is crucial for the viability of randomised controlled trials. In effect, such trials assume that the equation for the dependent variable under study will remain stable even when the equations governing the independent variables are altered. While this might often be true, it is by no means metaphysically guaranteed.

as a system of equations with exogenous independence. And this would then discredit the proposed reduction of causation (26), since that would imply that B and T must be causally unlinked, which by hypothesis is false.

We also find violations of (26) with joint normal probability distributions involving correlated 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.[19]

However, there is a way to exclude these unwanted cases and so uphold a version of the explicit reduction of causation (26) 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 further variables.

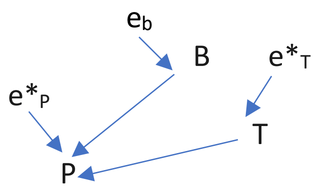
Consider what happens in the Hesslow example if we try to expand the equations (27) to include a variable for pregnancy P (which, remember, is correlated with both B and T). If we hold fixed the probabilistic independence of B and T, then such a system of equations with exogeneous independence throughout would need to take the form:

(27.1) $B = e_B$

(27.2) $T = e^*_T$

(28) $P = fB + gT + e^*_P$

[19] Note that the line that we get from linearly regressing X on Y is not the line that we get from linearly regressing Y on X. Under the assumptions of linear regression, these will be the same line only when X and Y are perfectly correlated.
In section 13 I showed that if we start with the equations for regressing Y on X
$X = e_X$
$Y= aX + e_Y$
and then rearrange terms to represent X as a function of Y rather than vice versa
$Y = e^*_Y$
$X= 1/aY+ e^*_x$
then the exogenous terms in the second pair of equations won't be probabilistically independent if those in the first are. Note how this all involved a single line ($Y = aX$, equivalently $X = 1/aY$) and different ways of recovering the joint distribution of X and Y from the way deviations about that line are conditionally distributed in the X and Y dimensions respectively.
By contrast, *different* lines are at issue when we linearly regress Y on X and X on Y respectively, and nothing rules out exogenous independence being satisfied by the structural equations associated with both sets of lines. And in fact we find this with multivariate normal distributions over correlated variables; indeed multivariate normal distributions over correlated variables can be characterised precisely as those for which exogenous independence is so multiply satisfied (Khatri and Rao 1976).



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 embeds the recursive structure of variables in the original equation set S.

The equations (27) displaying birth control pills B and thromboses P as independent are not so expandable, since any larger set of equations embedding (27) and also covering pregnancy P will violate the requirement of exogenous independence.

Let me now turn to mutually correlated variables in multivariate normal distributions. As I said, these allow a multiplicity of differently directed structural equations, not all of which can match actual causal structure. 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 over-supply of equations with exogenous independence in multivariate normal distributions as reflecting the way that the bridge principles sometimes fail to determine a unique causal order from a limited set of correlations. Earlier, in section 6, I pointed out that, given such underdetermination, a wider set of possible correlations can always suffice to fix a unique causal order for the variables at issue. If we now assume that such discriminating correlations involving further variables will always exist whenever 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 discriminating set of correlations. (If some other among the original set of equations were expandable, then the exogenous independence in the so-expanded equations would imply the existence of correlations that were inconsistent with the wider set of actual correlations that together with the bridge principles fix a unique casual order.)

Let me go a bit more slowly, and unpack this with the help of a concrete illustration. Suppose that some correlated X and Y have a bivariate normal distribution. Then both

(29.1) $X = e_X$
(29.2) $Y = aX + e_Y$

and

(30.1) $Y = e^*_Y$

(30.2) $X = bY + e^*_X$

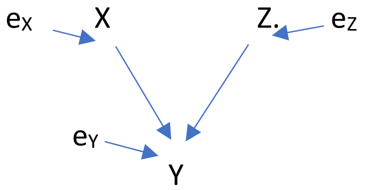
will be equation-DASs with exogenous independence that deliver the joint distribution of X and Y (where a and b are the standard regression coefficients).

But let us now suppose, in line with the idea that further correlations will generally determine which of these orderings reflect causal structure, that some further variable Z is correlated with Y, but not with X.

Then, I say, the only set of equations involving all of X, Y and Z that respects exogenous dependence will be the following set of equations, a set that embeds (29) rather than (30):

(31) $X = e_X$
$Z = e_Z$
$Y = aX + cZ + e_Y$

where a and c are again standard partial regression coefficients.



In support, consider what happens if we seek to expand (30) rather than (29) by regressing X on Y, and then Z on both Y and X:

(32) $Y = e^*_Y$
$X = bY + e^*_X$
$Z = fY + gX + e^*_Z$

with f and g again standard partial regression coefficients. If this set of equations satisfied exogenous independence, then we would expect to have a correlation between X and Z, which by hypothesis is not the case. Note that the regression coefficient g cannot be zero if X and Z are correlated with Y but not with each other ($g = r_{xz} – r_{xy}.r_{zy}$). So, if we if we put cancelling-out regression coefficients ($fg + b = 0$) to one side, the independence of the exogenous terms in (32) would force a non-zero correlation between X and Z[20]. Since it is given that this correlation is zero, the error terms can't be independent.

[20] There is one way in which the equations (32) could satisfy exogenous independence and yet respect the actual null correlation between X and Z—namely, if there were cancelling-out coefficients with $fg + b = 0$. We can deal with this possibility by refining the definition of expandability so as to rule out cases where an expanded set of equations respects the probability distribution only courtesy of cancelling out, while at the same time a different expanded set does this without relying on cancelling out. This would disqualify (32) as an expansion of (30), and so discredit (30)'s implied causal ordering of variables, since the different expansion (31) also respects the variables' probability distribution independently of any cancelling out.

This argument can be repeated for the other two possible expansions of (30) that also cover Z (namely, equations in which Z depends on Y, and X on Y and Z, and in which Y depends on Z, and X on Z and Y).

To return to the main line of argument, and in line with the above analysis of cases that cause trouble for the simple reduction (26), let me now hypothesise that the requirement of expandability will always be violated by systems of recursive equations with exogenous independence that do not correspond to 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:

(33) 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 an expandable set of deterministic equations. These specify how certain variables covary deterministically in a lawlike way.[21] 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 in the equations be probabilistically independent of each other and that this exogenous independence continues to be satisfied whenever those equations are expanded.

This might seem to rule out by fiat the possibility that (30) and (32) really do represent the real causal structure, with Y causing X, and both causing Z, and with the null X-Z correlation arising from a failure of faithfulness. But this possibility is not ruled out at all. What is ruled out is that (30) and (32) represent real causal structure with faithfulness failure while *at the same time* (31) displays exogenous Independence. And we have already been given reason to rule out this possibility, since the analysis of the equations (27)-(28) above showed that, in failure-of-faithfulness cases, exogenous independence will be violated when we attempt to expand causally misleading equations in which actual causes appear as dependent variables.

Think of it like this. In the current example, we have a probability distribution with X and Z correlated with Y, and X independent of Z, and cancelling-out regression coefficients. This set-up itself does not discriminate between these two causal possibilities: (a) X and Z are causes of Y, or (b) Y causes X, both cause Z, and the null X-Z correlation results from the cancelling out. Still, possible expansions do discriminate between these causal options, since in option (a) both expansions (31) and (32) will display exogenous independence, but in option (b) only (32) will. The refined definition of expandability is motivated by this kind of discrimination. The idea is that, when we have two incompatibly ordered expansions, both with exogenous independence, one of which needs cancelling out to respect the probability distribution, then we can be sure that the other one captures the real causal structure—after all, if the equations with cancelling-out did have the causal structure right, then they would be the only expansion which satisfied exogenous independence.

[21] 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.

16 Single-Case Causation

The current suggestion is that C causes E if and only if it is an ancestor of it in a recursive system of structural equations with exogenous independence. This now promises a better hold on single-case causation. (Did Joe Blogg's smoking actually cause his cancer?)

It will be useful to distinguish two issues here. First, what is it for one singular fact to be *actual caused* by another? Second, what is it for one singular fact to *counterfactually depend* on another?

It might not be immediately obvious why these need to be distinguished. However, it is a now familiar point that that there are cases of actual causation without counterfactual dependence, due to pre-emption, trumping, and so forth. Cases of this kind pose a central problem for David Lewis's programme of explaining actual causation in terms of counterfactuals.

To just give one illustration, consider this example, familiar from much recent literature in the Lewisian tradition. Suzy and Billy throw stones at a bottle. Suzy's arrives at the bottle first and shatters it. If her stone hadn't hit, Billy's would have shattered it anyway. This is a classic case where one event (Suzy's throw) *actually causes* another (bottle shattering) even though the latter *doesn't counterfactually depend* on the former.

Much recent work on both actual causation and counterfactual dependence appeals to "causal models" (Hitchcock 2018). These models posit directed relationships, in the form of "structural equations", and standardly pictorially represented 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 facts of actually causation and counterfactually dependence from these models.

The analysis of this paper complements this literature. While much progress has been made on the ways that causal models can help analyse actual causation and counterfactuals, there is no agreed view on what features of the real world these models represent. Far more attention has been paid to their use in evaluating single-case dependencies than in their content. In particular, there is no agreed account of what the directed "structural equations" mean. Those in the Lewisian tradition who invoke causal models generally regard these equations as encoding complex patterns of counterfactual dependence, and thus view the models a way of developing the programme of explaining actual causation in terms of counterfactuals. This leaves them, however, with the problem of explaining counterfactuals. Others, however, regard the equations as representing some species of primitive directed causal relations. This then allows them to view the "models" as a resource for analysing both actual causation and counterfactual dependence. But this advantage is bought at the expense of leaving the primitive causal relations unexplained. (Beebee and Menzies 2020.)

The approach I have adopted in this paper offers a way of transcending these options. I suggest that we should take there to be an arrow between two variables in a causal model just in case just in case one is a parent of the other in a system of deterministic equations with exogenous probabilistic independence. This then opens the prospect of explaining both actual causation and counterfactual dependence without having to leave the causal relations represented by the arrows as primitively unexplained.

It will be helpful at this stage to make some brief points about the different ways in which actual causation and counterfactual dependence can be analysed with the help of causal models.

At first pass, the recipe for reading counterfactual dependence off from a “causal model” is straightforward enough. Suppose the actual values of variables X and Y are m and n. Then “If X had been p, then Y would have been q”, for p ≠ m and q ≠ n, will be true just in case Y = q follows from the equations plus X = p and the actual values of all variables that are not descendants of X. (Galles and Pearl 1998, Hiddlestone 2005, Briggs 2012.)

Let me illustrate with the Suzy-Billy example and evaluating “If Suzy hadn’t thrown, the bottle wouldn’t have shattered”.

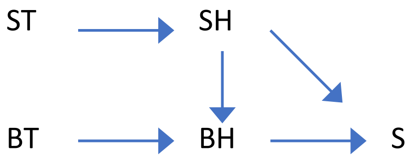
Let our “causal model” have the variables:

ST (Suzy’s throw), with possible values 1 (she throws) and 0 (she doesn’t)
BT (Billy’s throw), with possible values 1 (he throws) and 0 (he doesn’t)
SH (Suzy hits), with possible values 1 (she hits) and 0 (she doesn’t)
BH (Billy hits), with possible values 1 (he hits) and 0 (he doesn’t)
S (bottle shatters), with possible values 1 (it shatters) and 0 (it doesn’t).

And then we can specify these directed “equations” or dependencies:

ST: exogenous
BT: exogenous
SH = ST
BH = (min)(BT, |SH-1|)
S = (max)(SH, BH).

And we can pictorially represent these equations by:



In the actual world, the values of the variables are:

ST = 1
BT = 1
SH = 1
BH = 0

$S = 1$

Now, to evaluate "If Suzy hadn't thrown, the bottle wouldn't have shattered", we set $ST = 0$, leave $BT = 1$, and apply the equations, which then gives us $SH = 0$, $BH = 1$, and $S = 1$. The verdict is thus that the bottle would still have shattered, and so the claim "If Suzy hadn't thrown, the bottle wouldn't have shattered" is false.

It should be noted that this proposed recipe for evaluating counterfactual claims does not cover all counterfactuals. Most obviously, it doesn't determine the truth value of claims of the form "If X hadn't been m, then Y would have been p" in cases where X has a number of different alternative values to m and these imply different values for Y. In such cases we need some way of deciding what specific alternative value X would have had if it weren't m. Then there are also "backtracking" counterfactuals like "If I had jumped off the ledge, I would have been wearing a parachute", whose evaluation seems to call us to consider alternative causal ancestries for the antecedent variable as well as for the consequent variable.

Still, even if the analysis in terms of causal models does not deal with all kinds of counterfactuals, it still takes us a long way. As I said at the beginning of the paper, the Lewisian programme of analysing counterfactuals without invoking any causal information faces the problem that the evaluation of many counterfactuals seems to require us to hold fixed precisely those facts that are not *causally* dependent on the antecedent. The proposed recipe for reading counterfactual dependencies off from causal models offers a response to this challenge. It shows us how some specific counterfactual claims can be determined by causal facts, even if other counterfactual claims still call for a different treatment.

Let me now turn to actual causation. The general strategy for deciding whether some $X = m$ actually caused $Y = n$ is to see whether the latter counterfactually depends on the former *when certain "off-path" variables are held fixed at certain values*. The details then relate to the issues of which off-path variables should be set at which values.

To illustrate, let us return to Suzy and Billy once more. As I observed above, even though the shattering does not depend counterfactually on Suzy's throw, it was actually caused by it. In line with this, note how the shattering will counterfactually depend on Suzy's throwing if we *hold fixed the actual fact that Billy's stone didn't hit*. This vindicates the intuition that Suzy's throw was the actual cause. It also suggests the general hypothesis that $X = m$ actually causes $Y = n$ just in case there is a path of arrows from X to Y such that $Y = m$ counterfactually depends on $X = n$ when the values of all variables *off that path* are held fixed *at their actual values*.

However, while this formula works for Suzy and Billy, there are other cases where it seems to deliver the wrong answer. While there is general support for the idea that actual causation requires a path which yields counterfactual dependence when certain off-path variables are held fixed, cases different from Suzy and Billy's seem to call for fixing off-path variables at values other than their actual ones.

This is not the place to pursue this issue. I myself am attracted to an analysis, due to Brad Weslake (forthcoming), which in effect argues that the requirement of counterfactual dependence along a path is best understood in terms of chains of INUS conditionship (where an INUS condition is the familiar notion—insufficient but necessary part of an unnecessary but sufficient condition—originally due to J.L. Mackie). As Weslake sees it, a causal model implies that Ca actually caused Ea if the model shows that it was part of an INUS condition for part of an INUS condition for . . . Ea

The appeal to causal models to decide single-case issues leaves us with issues about the choice of models. It is one thing to read verdicts about actual causation and counterfactual dependence off from given causal models. It is another to decide what causal model should be applied to a given target situation in the first place. To adopt a model is effectively to place the target situation in a causal field, by implicitly holding the values of any variables omitted from the model fixed at their actual values. This raises a number of questions. Are there any objective principles pinning down which variables must be included in a model used to analyse any given target situation? If not, and the selection of variables is partly a matter of choice, how far will this lead to variant verdicts about actual causation and counterfactual dependence? And finally, if such verdicts are so variant, does this signify a deficiency in the approach in terms of causal models, or should we simply accept that facts of actual causation and counterfactual dependence are relative to choices of causal field?

17 Causation as a Guide to Action

Let me now return to generic causal claims of the form “C causes E” (“smoking causes cancer”, “birth control pills causes thromboses”, . . .). I now have argued that such claims should be understood as saying that C is an ancestor of E in a recursive system of structural equations with exogenous independence. And I have shown how we can infer such claims, reliably if not infallibly, from correlations by using the bridge principles.

As I have emphasised in this paper, much quantitative research in the non-experimental sciences is devoted to inferences of just this kind. What I haven’t discussed, however, is why we should be so interested in generic claims of the form “C causes E”. The points made in the last section might make one puzzled about this. We there saw how detailed knowledge of structural equations in the form of causal models can deliver conclusions about actual single-case causation and the nearby relation of single-case counterfactual dependence. But mere knowledge of generic claims of the form “C causes E” is by no means guaranteed to yield the detailed knowledge required to draw such inferences.

Indeed it will standardly leave us quite in the dark about such 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 decide these questions we’d need to know about what other variables are ancestral to cancer in a suitable set of structural equations, and about what values they had in Joe Blogg’s case, and just knowing that “smoking causes cancer” is likely to leave us highly ignorant about these matters.

The answer to this puzzle is that we are interested in generic claims like “smoking causes cancer” for a quite different reason than any concern with single-case dependencies. Even though they tell us little about single-case causal relations, generic casual claims can be a highly informative *guide to action*.

Note that statistical research will typically give us some numbers as well as the bare existence of a causal link. It will specify the strength of the correlation between cause and effect. It will tell us, say, that smoking S increases the probability of lung cancer L by a certain amount. Thus it might tell us that

(34) Prob(L/S) = 5%, where

Prob(L/not-S) = 1%.

(In truth, of course, different levels of smoking will make a different difference to the probabilities. But it will aid the exposition at this point to turn smoking into a dichotomous variable.)

We can view the correlation (34) as measuring how much the presence of S increases the probability of a sufficient condition for L. In the absence of S there is a sufficient condition for L in 1% of cases. But when S is present there is a sufficient condition in 5% of cases. Adding S to our situation means that it is 4% likelier we will have a sufficient condition for L. It shows that we have a 4% reason, so to speak, not to smoke if we want to avoid cancer.

In such a case, we needn’t know anything about the other factors relevant to L’s occurrence to know what difference S will make to the probability of L. Without knowledge of those other factors, it is true, we won’t necessarily know about actual causation and counterfactual dependence in particular cases, including in what will turn out to be our own case. But we don’t need to know such things to appreciate how much reason we have to quit smoking to avoid cancer.

It is important to note that it is the correlation between S and L *conditional* on any common causes of them both that we need to attend to in making decisions. Coarse-grained generalizations can certainly be suitable for guiding action even if they omit part of the effect’s full causes, but they do need to take into account any common causes of both S and L, for otherwise they will add to their indication of how often S contributes to a sufficient condition for L a spurious factor due to the S’s probabilistic association with other determiners of L.

Interestingly, a correlation between S and L (conditional on any joint causes) does not indicate how likely it is that L will actually be caused by S, but rather how likely it is to be *counterfactually dependent* on it. This is because any cases where L is actually caused by S but doesn’t counterfactually depend on it won’t contribute to the correlation between S and L. We would have had L anyway, even without S, due to the back-up cause, and so such cases won’t contribute to extra cases of L that are found when S in present. In the extreme case, if there were for some strange reason always a back-up cause for L whenever it was actually caused by S, then the probability of L given S would be no greater than without.

It might seem odd that correlations like (34) turn out not to be generalizations about actual causation, but rather about how often cancer counterfactually depends on smoking. But that is in fact in line with their role as a guide to action. If you are concerned to avoid cancer per se, it's the frequency of cases where cancer counterfactually depends on smoking that matters, not of those where it's actually caused by smoking. After all, you gain no advantage from not smoking in cases where you would have contracted cancer anyway, even if in those cases your smoking would have actual caused your cancer. (By way of analogy, note that Suzy has no reason to throw, if she wants the bottle to shatter per se, given that Billy's throw would shatter it anyway.)

There is much more to say about the connection between causation and rational action, and I hope to return to the topic at length in future work. Here let me content myself by making one general point. Some philosophers hold that causation cannot be analysed without appealing to the concept of *human action* or similar notions. The analysis of this paper argues that this attitude is mistaken. Causal facts are certainly relevant to rational action, for instance in the way just indicated. And perhaps the everyday *concept* of causation has important ties to notions of action. But at a metaphysical level it would be surprising if human action were prior to causation. After all, humans and their activities are part of the causal world, not prior to it. We should seek to understand causation first, without bringing humans into it, in the way I have done in this paper. This will then open the way to an understanding of humans and their activities as causal elements in a causal world.

18 "C causes E"

The points made in the last section give us reason to look a bit more closely at the meaning of generic causal claims like "C causes E". My proposed reductive analysis (33) implies that such claims are saying that C is an ancestor of E in an expandable system of structural equations with exogenous independence. And the last section explained that we are interested in quantitative versions of such claims because of the way they can tell us how much probabilistic reason we have to bring about the cause C in pursuit of the effect E.

But now we need to attend to the point that C can be an ancestor of E in a recursive equation system with exogenous independence, and yet the occurrence of C can render E no more likely that it would be in the absence of C. This is exactly what happens with failures of faithfulness. In Hesslow's example, for instance, birth control pills B are ancestral to thromboses T in a system of equations with exogenous independence, yet the probability of T is no higher given B than without B. Should we say C causes E in such cases? Do birth control pills B cause thromboses T?

The issue at this point becomes essentially terminological. I shall continue to use "C causes E" in line with my analysis (33), and so maintain that birth control pills causes thromboses even though they don't render them more likely. After all, this usage is in line with the original motivation for drawing attention to failures of faithfulness. Hesslow and others cited them precisely because they took them to offer counterexamples to the Unlinkage Condition, by showing us that an absence of correlation does not always signify an absence *of causation*. True, birth control pills causes thromboses through two different routes, with the probabilistic influence along one route being nullified by the influence of the other—but that doesn't mean, says Hesslow, and I agree, that they don't cause them at all.

At the same time, I am happy to grant that there is another generic sense of "C causes E" which does require that C makes E more likely. This sense goes with the thought that it is worth doing C in pursuit of E just in case "C causes E". Perhaps this is a more natural reading of "C causes E" than the one that requires only that C be ancestral to E in a system of equations with exogenous independence. Someone who says in English that "C causes E" is likely to be implying that it is rational to do C in pursuit of E. Still, as I said, this issue is now really terminological. We can simply specify how we want "C causes E" to be understood.

The divergence between my specified reading and one that requires C to make E more likely is not restricted to failures of faithfulness. There are also of course cases where C is ancestral to E in a system of equations with exogenous independence and yet C makes E *less likely*. There is nothing in the requirement of equational ancestry that demands a positive influence. Smoking is no doubt an equational ancestor of healthy lungs, in that the health of your lungs is a function inter alia of your smoking level, even though smoking of course makes healthy lungs less likely. I readily concede that it would be odd in normal English to say without qualification that "smoking causes healthy lungs", and to this extent grant that my reading of "C causes E" is something of a term of art. Still, it is the reading that is most convenient for our theoretical purposes.

We have now identified and distinguished a range of causal relations. My specified sense of "C causes E" signifies the crucial relation of ancestry in expandable equation systems with exogenous independence, and I have shown how this relation can be reliably though not infallibly evidenced by correlational patterns. I have also now allowed that "C causes E" in ordinary English might convey not only this ancestral relation, but also the requirement that C be positively correlated with E, or at least not be negatively correlated with it. We have also analysed and distinguished the associated single-case relations of actual causation and counterfactual dependence.

One prominent question in the philosophical literature on causation is whether causation is transitive or not. It would take us too far afield to address this issue here. But it is worth observing that there is no reason to expect a univocal answer. Some of the relations we have distinguished are transitive and some aren't. A satisfactory analysis of causal transitivity will need to start by being clear about what sort of causal relation is at issue.

19 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.[22]

[22] It is a moot point whether quantum mechanics is ultimately indeterministic. Collapse theories say so, but Everettianism or Bohmianism deny it. We can by-pass this issue here, however, as both Everettianism and Bohmianism still need to account for the *apparent* indeterminism that occurs when quantum

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 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 DNA 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

(35) $E = F(C, e_E)$

we will need

(36) $\text{Chance}(E) = F(C, e_{Ch(E)})$

If event E is undetermined until the time when it occurs, then the variables appearing on the right-hand side of the equation (36), including all influences packed into the exogenous variable $e_{Ch(E)}$, 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. If we were to allow that equations with the form of (36) could underpin causal relations, then we

superpositions interact with macroscopic systems. In line with this, I shall understand quantum indeterminism as covering whatever happens in such interactions.

would no longer be able to take for granted some of the links between correlations and causes that empirical researchers rely on. In particular, we would no longer be able to assume that correlation between pairs of causally linked variables will disappear when we control for the variables that mediate between them.

Recall the crucial earlier Determinism-Independence-Markov result (24) 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 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. For example, a source C might emit a spatially separated pair of electrons in the singlet state, with the spins of the electrons then being measured in given directions. Quantum mechanics predicts, and experiment confirms, that these measurement outcomes will be correlated in a way that is not screened off by any known features of the initial parent state, even if the background factors relevant to the two measurements are probabilistically independent. (Indeed the correlations displayed by measurements in different directions will have a collective structure that precludes their being screened off by *any* possible property of the initial state.)

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

(37) $\text{Chance}(E_1) = F(C, e_{Ch(E1)})$

(38) $\text{Chance}(E_2) = F(C, e_{Ch(E2)})$

And now the relevant point is that the non-determinism of these equations *leaves room* for the outcomes to co-ordinate themselves within the freedom, so to speak, left open by their non-determination. 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 (24). 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 shows that, once this room is made available, nature sometimes makes use of it.[23]

[23] From this perspective, the EPR quantum correlations are arguably less surprising than they first appear. Think of it like this. Determinism with exogenous independence leaves no possibility of correlations that aren't screened off by common antecedents. On the other hand, in the absence of this underlying deterministic structure, as in the EPR set-ups, then there is probabilistic room, so to speak, for results to become correlated in a way that isn't screened off by any feature of their common source. And it turns

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

(36) $\text{Chance}(E) = F(C, e_{Ch(E)})$

as

(39) $E = F(C, e_{Ch(E)}, e_{E)})$

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 (39) as "pseudo-deterministic". Dependencies that appear probabilistic only because they omit the totality of determining factors are often termed "pseudo-*in*deterministic". But contrast, while equations like (39) have the appearance of determining the effect variable E, this conceals the way that the e_E term on the right-hand side is 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. And this means that we can happily relax our analysis of causation to allow systems of pseudo-deterministic structural equations to ground causal relations in the same way that deterministic structural equations do. Because the independent exogenous variables function similarly in both deterministic and pseudo-deterministic equations, this relaxation will allow us to 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 the Unlinkage Principles can be expected to hold except in special cases of faithfulness failure.

out that, as soon as nature has this room, it uses it to produce unscreenoffable correlations. What's so surprising about that?

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

(40) X causes Y if and only if it is an ancestor of Y in an expandable recursive structure of deterministic *or pseudo-deterministic* equations with independent exogenous terms.[24]

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 pseudo-deterministic. If we try to put these equations into the pseudo-deterministic form (39), the "dummy" exogenous variables representing the undetermined manifestation of the outcomes on the two wings will not come out as 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 Causal Markov Condition and its implied 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 Causal Markov Condition is no problem for practical non-experimental 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 Causal Markov Condition and its Linkage corollaries in inferring causes from correlations.

Still, one might wonder where the EPR correlations leave my claim that the Causal Markov Condition is a deductive consequence of my proposed analysis of causation. If my analysis

[24] My earlier analyses of 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.

does indeed imply the Causal Markov Condition, and the EPR correlations show the Causal Markov Condition is 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 recursive structures of deterministic or pseudo-deterministic 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 pseudo-deterministic equations with probabilistically independent exogenous variables.

This does now mean, however, that my original Conditional and Unconditional Linkage Principles (2) and (3) 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 non-experimental researchers, but it is needed if we want to keep the logic straight.[25]

20 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 recursive systems of deterministic or pseudo-deterministic equations with probabilistically independent exogenous terms. Each dependent variable in such an equation is causally posterior to its exogenous variable and other independent variables. Those independent variables will in turn have their own equations displaying them as causally posterior to their own exogenous and independent variables. The eventual upshot will be that every variable is causally descended in a certain way from the set of exogenous variables.

Now in actuality, 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 structural equations with

[25] Did I not argue earlier in footnote 8 that the Causal Markov Condition implied the original Linkage Principles (2) and (3) without qualification? 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 systems of equations with exogenous independence, the Causal Markov Condition and hence the Linkage Principles will no longer apply to EPR quantum correlations.

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 structural equations are the result of quantum superpositions resolving themselves into determinate outcomes when they interact with macroscopic systems. (Sometimes, I have suggested, this determinacy will temporally precede the relevant dependent variable, in which case we will have a deterministic structural equation. In other cases, the relevant structural equation will be only pseudo-deterministic, with a "dummy" exogenous variable signifying that the dependent variable is undetermined until it occurs.)

One immediate consequence of viewing the exogenous variables in this way is that we can attribute their probabilistic independence to the typical unconnectedness of the processes giving rise to quantum "collapses". In general, distinct quantum collapses will occur in interaction with unrelated macroscopic systems. It is true that, in the special circumstances of the EPR-type experiments, where the measurements on the two wings are carefully arranged, we will find co-ordinated collapses—which was why the "dummy" exogenous variables in my formalization of the EPR set-up violated the requirement of probabilistically independence. But I take cases like this to be the exception rather than the rule. In the wild, the different macroscopic systems with which spatially distributed quantum superpositions interact will not be co-ordinated in the way required to display EPR-like correlations, and we can thus take it that any "collapses" they prompt will display probabilistic independence.

Explaining the probabilistic independence of exogenous variables is one thing. But our current subject is 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.

Does this now explain why quantum collapses always temporally precede the further events that depend on them in systems of recursive equations? I am not sure. At first pass it might seem natural to hold that an event that only becomes determinate at a certain time cannot influence goings-on at earlier times. But on reflection this simple thought arguably begs the question. Once a quantum outcome has become determinate, what rules out its being related by structural equations to events at earlier times, except some prior assumption that such temporally reversed influence is impossible?

At this stage I think it matters how we analyse quantum "collapses". In particular, I think that the Everettian interpretation of quantum mechanics has the resources to block this charge of begging the question. On non-Everettian interpretations, quantum "measurements" have unique outcomes, displayed in in the one actual universe, whch arguably means that non-Everettian views can offer no principled reason why later quantum outcomes should not influence earlier events, apart from the question-begging assumption that such temporally reversed influences are impossible. But on the Everettian view, quantum measurements result in all possible outcomes, each displayed in its own branch of reality, with the universe repeatedly splitting as time progresses. This allows Everettians to offer a rationale for rejecting temporally reversed influences. On their branching picture of reality, a chancy quantum outcome is not a determinate element of the world that preceded the apparent "collapse". That world evolved into multiple futures, each with its own determinate outcome. And this in itself would seem to rule out any given quantum outcome influencing pre-split events. Why could possible privilege that outcome as an influence on earlier events, when all the alternative possible outcomes are equally part of those events' futures?

On the Everettian picture, then, we seem to have a good explanation for why chancy quantum outcomes cannot be related by structural equations to events that precede them in time. The world inhabited by those earlier events doesn't contain those quantum outcomes as determinate elements, and so laws coordinating events in that world cannot portray those earlier events as response to those quantum outcomes. It is only once those outcomes have become determinate, each on their own future branch of reality, that they are available, so to speak, to exert their own distinctive influence on what later happens in that branch.

My analysis of causation thus makes the temporal asymmetry of causation a consequence of the temporal asymmetry of Everettian branching. The quantum events that provide the independent exogenous terms in structural equations are also responsible for the constant branching of the universe, as each possible outcome in any apparent quantum "collapse" instigates its own branch of reality. Because these outcomes are peculiar to their own branches, they can only be related by law to later events within those branches, not to earlier pre-split events. The asymmetry of causation thus turns out to be a natural consequence of the asymmetric orientation of quantum branching in time.

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. However, the Lewisian tradition simply assumes this asymmetry without further explanation. From my perspective, this puts the cart before the horse. The "asymmetry of overdetermination" is a real enough phenomenon, but it is not prior to the asymmetry of causation. As I see it, the Lewisian 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 correlated with each other. Two variables that are both correlated with a common ancestor in a system of deterministic or pseudo-deterministic equations with exogenous independence will necessarily be probabilistically dependent. By contrast, nothing in my account requires two variables with a joint descendant to be correlated. These points mean 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 typically 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. From my perspective, then, the "asymmetry of overdetermination" is not an independent phenomenon to be invoked to explain the asymmetry of causation, but itself an upshot of the way that causation is itself asymmetrically orientated in time.[26]

21 The Ubiquity of Chance

I have argued that the causal structure of the world arises from the way certain variables are governed by deterministic or pseudo-deterministic structures of equations with exogenous independence. This analysis allows us to understand, first, how causes can be inferred from correlations, second, what grounds actual causation and counterfactual dependence, and finally why causation is temporally asymmetric.

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 *idea* 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 analysis does not derive from the concept of causation, I am happy to allow that we can coherently apply the 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

[26] 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.

truth, causation depends on 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 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. Of course some prior circumstances do make others overwhelmingly likely. 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 result. Variables representing those quantum 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. 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 presenting the effect, which will then constitute those factors as its causes. I take the underlying quantum nature of reality to give us every reason to accept this thesis.[27]

References

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

Blalock, H. 1967 *Causal Inferences in Nonexperimental Research* New York: W. W. Norton

Blanchard, T. and Schaffer, J. 2017 “Cause without Default” in Beebee, H., Hitchcock, C. and Price, H. eds *Making a Difference*, Oxford: Oxford University Press 175–214

Briggs, R. 2012 “Interventionist Counterfactuals” *Philosophical Studies* 160: 139–66.

Cartwright, N. 1979 “Causal Laws and Effective Strategies” *Noûs*, 13: 419–437

[27] I would like to thank Thomas Blanchard, Julien Dutant, Toby Friend, Totte Harinen, Jenn McDonald, Davide Pigoli and Henry Taylor for comments on earlier versions of this paper.

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

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

Dearden, L. Ferri, J and Meghir, C. 2002 "The Effect of School Quality on Educational Attainment and Wages" *Review of Economics and Statistics* 4: 1-20

Dupré, J. 1984 "Probabilistic Causality Emancipated" in French, P., Uehling, T. and Wettstein, H. eds *Midwest Studies in Philosophy IX* Minneapolis: University of Minnesota Press 169–75

Durkheim, E. 1897 *Suicide: A Study in Sociology* translated by Spaulding, J. and Simpson, G. London: Routledge & Kegan Paul

Ehring, D. 1987 "Papineau on Causal Asymmetry" *British Journal for the Philosophy of Science* 38: 81-7

Einstein, A., Podolsky, B. and Rosen, N. 1935 "Can Quantum-Mechanical Description of Physical Reality be Considered Complete?" *Physical Review* 47: 777–780

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

Fisher, R. A. 1925 *Statistical Methods for Research Workers* Edinburgh: Oliver and Boyd

Galles, D. and Pearl, J. 1998 "An Axiomatic Characterization of Causal Counterfactuals" *Foundations of Science* 3: 151–82

Good, I. J. 1961-2 "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

Halpern, J. 2016 *Actual Causality* Cambridge Mass: MIT Press

Healey, R. 1997 "Nonlocality and the Aharonov-Bohm Effect" *Philosophy of Science* 64: 18-41

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

Hausman, D. and Woodward, J. 1999 "Independence, Invariance and the Causal Markov Condition" *The 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

Hiddleston, E. 2005 A Causal Theory of Counterfactuals” *Nous* 39: 232–57

Hitchcock, C. 2001 “The Intransitivity of Causation Revealed in Equations and Graphs” *Journal of Philosophy*, 98: 273–99

Hitchcock, C. 2018 “Causal Models” in Zalta, E. ed *The Stanford Encyclopedia of Philosophy* (Summer 2018 Edition)

Hofer-Szabó G., Rédei, M., and Szabó, L. 2013 *The Principle of Common Cause* Cambridge: Cambridge University Press

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

Khatri, C. and Rao, C. 1976 “Characterizations of Multivariate Normality. I. Through Independence of Some Statistics” *Journal of Multivariate Analysis* 6: 81-94

Lazarsfeld, P. and Rosenberg, M. 1955. *The Language of Social Research: A Reader in the Methodology of the Social Sciences* Glencoe Ill: Free Press

Lewis, D. 1973 “Causation” *Journal of Philosophy* 70 :556-567

Lewis, D. 1979 “Counterfactual Dependence and Time’s Arrow” *Nous* 13: 455-76

Loewer, B. 2007 “Counterfactuals and the Second Law” in Price, H. and Corry, R. eds *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited* New York: Oxford University Press, pp. 293-326

Papineau, D. 1988 "Response to Ehring’s ‘Papineau on Causal Asymmetry’” *British Journal for the Philosophy of Science* 39: 521-5

Papineau, D. 1991 "Correlations and Causes” *British Journal for the Philosophy of Science* 42: 397-412

Papineau, D. 1992 “Can We Reduce Causal Direction to Probabilities?" *Philosophy of Science Association Vol 1992* 2: 238-52

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

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

Pearl, J. 2017 “The Eight Pillars of Causal Wisdom” edited transcript of a lecture at West Coast Experiments April 2017 https://ftp.cs.ucla.edu/pub/stat_ser/wce-2017.pdf

Pearl, J. 2018 *The Book of Why* London: Allen Lane

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

Pollock, S. 2014 “Econometrics: A Historical Guide for the Uninitiated” *Working Paper 14/05* Department of Economics, University of Leicester

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

Schaffer, J. 2004 “Counterfactuals, Causal Independence and Conceptual Circularity” Analysis 64: 299-309

Schurz, G. 2017 “Interactive Causes: Revising the Markov Condition” *Philosophy of Science* 84: 456–79

Schurz, G. and Gebharter, A. 2016 “Causality as a Theoretical Concept” *Synthese* 193: 1073-103

Simon, H. 1953 “Causal Ordering and Identifiability” in Hood, W. and Koopmans, T. eds *Studies in Econometric Method: Cowles Commission for Research in Economics* 49-74

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

Sosa, E. and Tooley, M. 1993 “introduction” in Sosa, E. and Tooley, M. eds *Causation* Oxford: Oxford University Press 1-32

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

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

Strevens, M. 2007 “Essay review of Woodward *Making Things Happen*” *Philosophy and Phenomenological Research* 74: 233-49

Strevens, M. 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

Wenglinksky, H. 2007 *Are Private High Schools Better Academically than Public High Schools?* Washington DC: Center for Education Policy

Weslake, B. forthcoming “A Partial Theory of Actual Causation” *British Journal for the Philosophy of Science*

Woodward, J. 2003 *Making Things Happen*

Woodward, J. 2008 “Response to Strevens” *Philosophy and Phenomenological Research* 77: 193-212

Woodward, J. 2016 "Causation and Manipulability" in Zalta, E. ed *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition)

Wright, S. 1921 "Correlation and Causation" *Journal of Agricultural Research* 20: 557–85

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