The Statistical Nature of Causation

David Papineau

1. Introduction

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.[1]

This poses an obvious philosophical challenge. Why do these techniques work? Why do causal structures have a distinctive correlational signature?

Yet none of the major metaphysical approaches to causation offers any answer to this challenge. As far as I know, no philosophers developing counterfactual theories of causation, or dispositional theories, or regularity theories, or process theories, so much as raise the issue.

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"[2]. It is creditable 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 poses a manifest challenge to metaphysicians. Can it just be a coincidence that the casual structures line up so nicely with the correlational patterns?

This paper will show that this match is not a coincidence. The statistical signatures that guide the non-experimental scientists are built into the nature of causation. I shall offer a reductive analysis of directed causal influence that explains the success of the non-experimental techniques.

In the middle of the last century a number of philosophers, including Hans Reichenbach (1956), I.J. Good (1961-2) and Patrick Suppes (1970), sought to explain the connection between causes and correlations by "probabilistic theories" of causation which aimed to reduce the former directly to the latter. That approach, however, runs into well-known problems. The reduction I shall offer will be different. I shall avoid the problems facing probabilistic theories by reducing causation to underlying structural equations with independent error terms rather than directly to surface correlations. The moral of the failure of probabilistic theories is not to ignore the connection between causes and correlations, as the philosophical mainstream has done, but to seek a better explanation.

---

[1] This 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 (see Pollock 2014), including Nobel prize winners Jan Tinbergen, Ragnar Frisch and H.A. Simon, and by social scientists, including Paul Lazarfeld (Lazarfeld and Rosenberg 1955) and H.M. Blalock (1972). 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.

[2] Pearl 2017 9, 2018 116.

The general plan of the paper is as follows.

Sections 2-6 outline the standard procedures for inferring causes from correlational patterns.

Sections 7-10 consider the possibility of reducing causation directly to correlational patterns and explain why this faces problems.

Sections 11-19 argue that a different reduction can also explain the correlational techniques while avoiding the problems: the key is to reduce causation, not directly to observed correlations, but to underlying systems of structural equations with probabilistically independent exogenous variables.

2. 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 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.

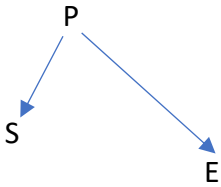
This further evidence now argues that school funding does not cause higher examination results after all, but that their initial correlation was rather due to their both having a common cause, higher parental income. The children in well-funded schools do better, not because of the schools, but because 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.[3]

The overall evidence thus suggests the following casual *directed acyclic structure*[4] ("DAS" henceforth):

[3] 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.

[4] The more familiar coinage is directed acyclic "graph" (DAG). I have adopted "structure" instead to stress that my concern is with worldly relationships between worldly quantities, and not with the means by which we might represent these relationships.

(1)



In such a causal DAS, an arrow means that the variable at the head of the arrow causally influences 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] The notion of "causal influence" needs further unpacking, but for now it will be convenient to take it as read; by the end of the paper I will be in a position to explain it.

This educational example of an inference from correlations to causal structure illustrates the philosophical challenge I wish 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.

3. 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 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.

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

[5] "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}, . . .

separate probabilities. 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.)

4. 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 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?

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.[8] In this section and the next two, I shall articulate these

[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?)

[8] The two classic sources are Spirtes et al 1993 and Pearl 2000.

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.

Given this notion, we can immediately state two principles connecting causes and correlations.

First, a *Linkage Condition*:

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

Second, a *Conditional Linkage Condition*:

(3) If two 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 conditions are normally derived from a prior "Causal Markov Condition", but for our purposes it will be more perspicuous simply to focus directly on these consequences.[9]

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 Condition*:

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

---

[9] 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 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 6 below). So understood, and supposing there are no further requirements on causal structures beyond these (see section 18 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.

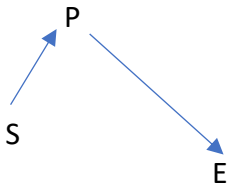
Second, a *Conditional Unlinkage Condition*:

(6) If two correlated variables are screened off by other variables {X}, then they are not causally linked by any chains of variables that do not contain any of {X}.[10]

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.

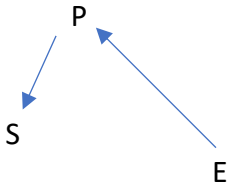
This now further narrows down the causal possibilities. True, this does not yet uniquely determine the structure (1), with P as the common cause of S and E, that I initially suggested as the natural causal interpretation of the correlations. For, even given the four posited principles, the correlations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

(8)



and

(9)



Still, these three structures are the only options.

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:

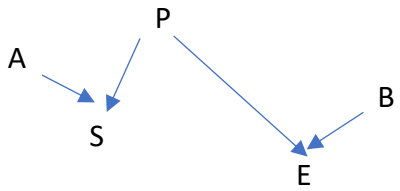
---

[10] The Faithfulness Condition can be stated as:

(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.)

(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.

5. The Power of the Bridge Principles

Let me call conditions (2), (3), (5) and (6)—equivalently the Causal Markov and Faithfulness Conditions—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 their 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.[11]

Of course, empirical researchers don't always need to infer their causal conclusions from correlations alone. In practice they will standardly help themselves to partial prior causal knowledge to narrow down the causal possibilities and simplify their inferential task. 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. But this kind of assistance from common sense or temporal ordering is by no means essential.

6. Including Common Causes

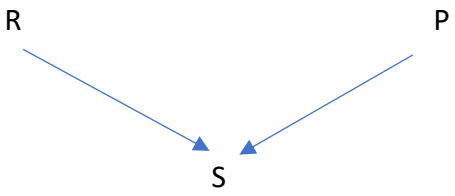
So the bridge principles allow us to infer unique causal structures from sufficiently rich sets of conditional and unconditional correlations. But there is an immediate worry about taking this to show that they can uncover genuine causal relationships. Our examples so far, and the general theorem mentioned 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 our focus here is on drawing causal conclusions about *reality*, not causal conclusions *relative to some model*.

[11] Theorem 4.6 Spirtes et al 1993 94.

So we face an obvious question. What guarantees that causal conclusions derived from correlations among some limited set of variables will not be undermined if we expand this set to include further variables?

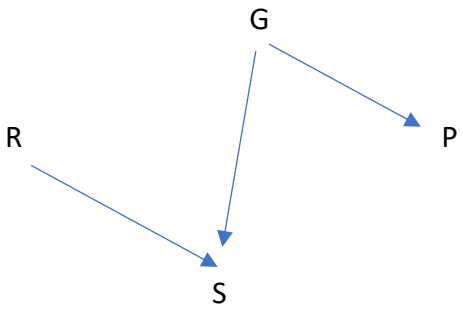
This is by no means an idle worry. Imagine that, in our original example, family religion (R) is also correlated with school type S, but is uncorrelated with parental income P. Then the bridge principles would dictate:

(11)



But now suppose that P does not in fact cause S but that instead they are joint effects of a common cause (perhaps, not entirely implausibly, they are both effects of grandparental income G). Then the true causal DAS would be:

(12)



And this case basing our causal analysis on the correlations among the original limited set of variables would have led to an erroneous conclusion about causal structure. We would have concluded parental income P causes school type S, when in truth it doesn't.

This example shows how the application of the bridge principles to coarse-grained sets of variables presupposes that any correlation between some X and Y that is not screened off by any of those variables signifies a *direct* causal link. This presupposition then implies that any unscreened-off correlation between some X and Y won't be due to their being indirectly linked via a common cause—rather X must directly cause Y, or Y must directly cause X. And of course this implication can turn out to be mistaken when we bring in further variables, as just illustrated by the example involving grandparental income G.

Still, this danger is importantly limited. Coarse-graining will only lead us substantially astray if it omits *common causes* of included variables. Bringing in extra variables can show that what initially appeared as direct links between some X and Y are in fact indirect links. But the initial appearance will only have been significantly misleading if X and Y turn out be indirectly linked via a common cause. In all other cases all that we will have discovered is that in truth X causes Y, or Y causes X, not directly, but via intermediaries. And of course this

will be no surprise. It is only to be expected that causal influences that appear direct "relative to some model" are in reality mediated.

Let me be more precise. Suppose that we are given some limited set of variables {V} the correlations between suffice for the bridge principles to fix a causal order among them.

A first point to note is that bringing in extra variables beyond {V} will not in itself change the unconditional and conditional correlations that obtain between the variables in {V}. These are fixed features of {V}, so to speak, and will remain stable as we add further detail. For example, adding further variables to our educational analysis won't alter the fact that school type is unconditionally correlated with examination results.

At most, bringing in further variables will show us that what looked like direct unscreened-off correlations between variables X and Y in {V} are in truth screened off by some set of variables {Z} outside {V}—as in our example above grandparental income G turned out to be screened off the previously unmediated correlation between P and S.

Still—on the assumption that the full underlying reality, so to speak, satisfies the bridge principles—this can only happen in two ways. Either the new variables within {Z} are causally intermediate between the originally directly correlated variables X and Y, or they are common causes of them. (The Conditional Unlinkage Condition (6) implied that a correlation between two variables is screened off only if we control an intermediary from each causal chain linking them.)

Now, as observed above, in the first case of causal intermediaries the extra fine-graining will scarcely have *overturned* any causal conclusions derived from the original coarse-grained {V}. We will simply have confirmed, as we would always have assumed, that a causal influence that appears direct relative to {V} is in reality mediated by further intervening variables.

It is only in the second case, where new screening-off variables are common causes, that the causal conclusions derived from the original coarse-grained set are seriously threatened. For then, as in our grandparental income G example, the screeners-off can show that X doesn't cause Y, but that they are joint effects of a common cause.

The moral is that, if the bridge principles are satisfied by the fully fine-grained reality, then their application to more coarse-grained sets of variables extracted therefrom can also be relied on, subject only to the proviso that those sets do no omit common causes of variables they do include.[12]

This metaphysical point raises obvious epistemological questions. Will empirical researchers ever know that they have included enough common causes in their analysis? This is of

[12] I have been speaking of "full reality itself satisfying the bridge principles". This is best understood in terms of a level of variable inclusion at which the bridge principles are satisfied and beyond which they stay satisfied. This understanding 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 one level of analysis always becomes indirect at finer levels.

course a real worry. Practical researchers seeking to derive causal conclusions from 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.[13] 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.

7. 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 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 always hold. 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 have at least some validity.

This now returns us to the metaphysical challenge I started with. Why does causation display itself in correlational signatures?

One option here would be to view the bridge principles as contingent truths. 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 Judea Pearl's attitude. "A gift from the gods." As he sees it, 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

[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 metaphysics of randomized trials see footnote 21 below.

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.

8. 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 correlations that omit no common causes can fix 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.

From this perspective, the bridge principles would no longer be contingent, but metaphysically necessary. They would be simply fall out of the way causal structures are constituted by correlational ones.

The original "probabilistic theories" of causation due to Hans Reichenbach (1956) and I.J. Good (1961-2) and Patrick Suppes (1970) were all variations on the central idea that:

(13) 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.

This formulation, however, was ill-equipped to deal with structures involving a multiplicity of variables. (Moreover, by appealing to temporal order, it blocked the prospect of grounding the earlier-later asymmetry in causal asymmetry.)

The approach currently being suggested avoids these drawbacks. The approach to causal structure via the bridge principles can deal with structures of any complexity and makes no appeal to temporal information. True, I have as yet offered no explicit reduction of the form *A causes B if and only such-and-such*, as opposed to hypothesizing that causal structures are nothing over and above the correlational structures from which they can be inferred via the bridge principles. Still, if such an explicit reduction is wanted, one is not too far away. The bridge principles (plus one extra assumption) imply this necessary and sufficient condition for causation:

(14) A causes B if and only if A is correlated with B, and everything correlated with A is correlated with B, and something correlated with B is not correlated with A.[14]

The basic idea here is that effects are distinguished from causes by having some independent sources of variation. Intuitively, the things that co-vary with any given cause A—its causes and its effects—will be correlated with its effect B—since they will either cause B or share a common cause with it; but some things will be correlated with B—its other causes—that won't be correlated with A. (The extra assumption needed to establish (14), in addition to the bridge principles, is that there will always *be* such independent sources of variation.)

Attractive as this neo-probabilistic reduction of causation might appear at first sight, we shall now see that it is flawed as a metaphysical analysis of causation. It locks onto the symptoms of causation, rather than its underlying nature.

9. The Bridge Principles Examined

The neo-probabilistic reduction of causation takes the bridge principles to be metaphysically necessary. But the prior question is whether they are even generally true.[15] The literature contains challenges to all the bridge principles. Some of these are relatively superficial, and can be parried by the neo-probabilistic theory of causation just outlined. But objections to the Faithfulness Condition cannot be so easily dealt with, and will require us to move beyond the neo-probabilistic account.

At this stage, I shall put to one side non-local quantum correlations like those between measurements on spacelike separated entangled particles ("EPR" correlations henceforth, after Einstein, Podolsky and Rosen 1935). Correlations like these have a non-standard structure (they can't possibly be screened off by any features of their common source) which differentiates them from more familiar everyday correlations. I shall discuss the relevance of non-local quantum correlations to causal structure in sections 17 and 18 below.

Let me start with the two conditions that allow us to infer causal links from correlations. First is the simple Linkage Condition (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.

---

[14] This formulation is inspired by the theory defended in Daniel Hausman's rich and penetrating *Causal Asymmetries* 1998. It should be noted that Hausman does not himself aim to reduce causation to correlations, but rather to a modally more robust relation of "causal connection" that he invokes to circumvent failures of faithfulness.

[15] A different kind of objection to probabilistic theories of causation relates to the way in which they average over inhomogenous single-case chances and so cannot straightforwardly deliver conclusions about singular or "token" causes. See for example Cartwright 1979 Dupré 1984. Issue of singular causation lie beyond this paper (though see the brief remarks in my final section), but I agree entirely that traditional probabilistic theories are ill-suited to deal with them.

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 properties involved each have systematically connected values.

I now turn to the Conditional Linkage Condition (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.

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 everyday examples is due to the common cause being under-described. If the precise angle of the impact were given, then this would render each ball's trajectory probabilistically irrelevant to the other's. Proposed counter-examples like Salmon's can thus be dealt with as violating the requirement that we are dealing with variables that do not omit any common causes.

10. The Failure of Faithfulness

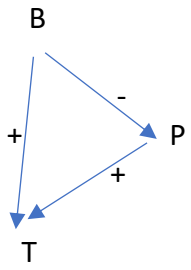
I turn now to objections to the Faithfulness Condition. For simplicity I shall focus on the simple Unlinkage Condition (the same issues arise with the Conditional Unlinkage Condition). The Unlinkage Condition said:

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

This claim faces a difficulty that cannot be easily dismissed. Imagine that one variable causes 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 (15) 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.

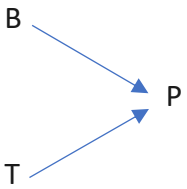
(15)



Clearly this kind of example can easily be multiplied. (Imagine that high parental income actually has a *negative* influence on examination results, but that this is cancelled out by the direct positive influence of well-funded schools . . .)

Such "failures of faithfulness" 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:

(16)



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.[16] 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.

11. Linear Regression

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 deterministic *structural equations* assumed by such traditional statistical methods as analysis of variance, regression analysis, and combinations thereof. I shall now offer a reductive analysis of causation in terms of such structural equations.

[16] 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.

This analysis will still uphold the bridge principles as the basis on which causal conclusions can be inferred from correlations. But now only the Linkage and Conditional Linkage Conditions will be delivered as metaphysically necessary consequences of the analysis. By contrast, the Unlinkage and Conditional Unlinkage consequences of the Faithfulness Condition will come out as principles that can generally be relied upon but are by no means metaphysically guaranteed.

Let me approach my reduction by running through 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:

(17.1) $P = e_P$

(17.2) $S = aP + e_S$

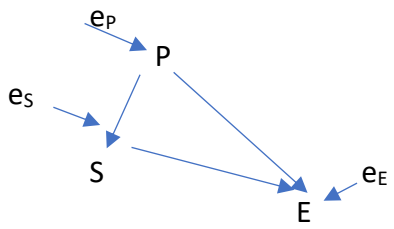
(17.3) $E = bP + cS + e_E$

(To repeat: P = parental income; S = school funding; E =examination results)[17]

The equations represent deterministic relationships. The subscripted rightmost e-terms are called "error terms" and 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. (Note that this is a different kind of DAS--an *equation-DAS* I shall call it henceforth, by contrast with our earlier *cause-DAS*s.)

(18)



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

[17] I shall 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", because this obscures the way in which the non-standardized slopes, and more generally the forms, of structural equations are more robust that the variances of their variables, as explained in section 13 below.

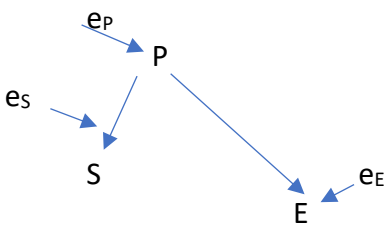
In our example, we are supposing that examination results E don't co-vary at all with schooling S once parental income P is held constant. So then the regression coefficient c will be zero, and the equations will have the simpler structure:

(19.1) $P = e_P$

(19.2) $S = aP + e_S$

(19.3) $E = bP + e_E$

(20)



Now, the recursive structures of equation sets like (17) and (19) strongly invite a *causal* interpretation. It is natural to 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 (17) and (19) are really *equations*, what is to stop us presenting them in a transformed order? For example, viewed purely as a set of equations, (20) could happily be rewritten as:

(21.1) $S = e^*_S$

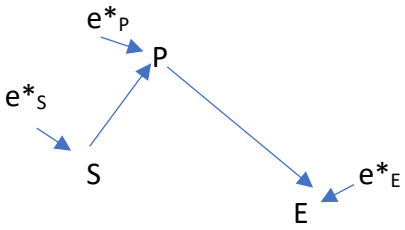
(21.2) $P = 1/aS + e^*_P$

(21.3) $E = bP + e_E$

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

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

(22)



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

So—what tells us to do things the first way rather than the second one? 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 viewing things. 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 independent information about possible causal ordering.

11. Independent Error Terms

The crucial factor is whether the *error terms* are *probabilistically independent*. The regression analysis approach to causation rests on the assumption that we can read causal structure off from the equations *only if the error terms are not correlated with each other*. This then provides a rationale for taking the first set of equations rather than the second to represent casual structure. If the error terms e_S, e_P, e_E are probabilistically independent, then this implies that the first version has the causal structure right. And by the same coin, this means that the second must get the causal structure wrong, since the error term e^*_S in (21.1) is a linear function of the error term e_E in (21.3) plus another independent term, and so cannot be probabilistically independent of it.

This independence requirement isn't just an arbitrary add-on to regression analysis. It is integral to the way such equations are constructed and used for prediction and explanation.

To see this, go back to our first two regression equations above:

(17.1) $P = e_P$

(17.2) $S = aP + e_S$

If the error terms e_P and e_S are independent, then these equations mean that the values of P are determined by one set of factors, e_P, and the values of S in turn are fixed by P *and* an independent set of factors e_S. This is why we can use the value of P alone to predict the mean value of S—the variations in S for any value of P are due to factors independent of P's value.

Note how such prediction is not possible with the "transformed" equations

(21.1) $S = e^*_S$

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

We can't use these equations in reverse to predict an expected value for P from a given value for S. Now that e^*_P is probabilistically associated with S—recall that $e^*_P = -e_S/a$—we can no longer view it as an independent "noise" term added to the influence that S has on P. (Rather it is a sort of correction term that subtracts from the contribution of S the part due to influences on S other than P; the barrier to prediction is then that S's value on its own, without being given the probability distribution of P, is uninformative about the nature of this correction.)[18]

12. Causal Structure and Error Term Independence

The independence of error terms in recursive systems of structural equations holds the key to causal structure. Such independence means that the values of dependent variables are the upshot of influences that can be factorised into independent sources. Since the error terms are probabilistically independent of each other, since they display no systematic tendency to vary in concert, they are constituted as causally unlinked. The dependent variables, by contrast, are built up from these factorizable influences, and are thereby constituted as their effects.

So far I have illustrated the idea in a maximally simple case with one explicit cause and an independent error term. But the idea that independent error terms constitute causal structure applies more generally, and in particular to effects of a plurality of independent variables than can themselves be correlated with each other. 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 and associated equation-DAS would then be the earlier (17) and (18):

(17.1) $P = e_P$

(17.2) $S = aP + e_S$

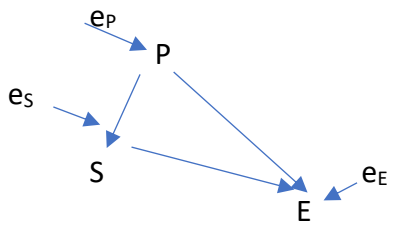
(17.3) $E = bP + cS + e_E$

[18] If someone did seriously want to use regression analysis to predict P on the basis of S, then they would not use the transformed equations (21.1, 21.2), but rather the equations that result from "regressing P on S".

(23.1) $S = e^*_S$
(23.2) $P = dS + e^{**}_P$

Note that the line P = dS that we get from regressing P on S is not the line S = aP that we get from regressing S on P. They will be the same only when P and S are perfectly correlated. Using the former line to predict mean values for P from S would require the error term in (23.2) to be probabilistically independent of values of S, to justify factorising the influences on P into two independent elements. In the case at hand, however, we would not expect this assumption to be satisfied. If e_S is probabilistically independent of P in (17.2) then it can't also be that e^{**}_P is independent of S in (23.2). This term e^{**}_P will be positively correlated with S, with the result that the "predictions" derived from (23.2) will tend to underestimate the extent to which P values deviate from their average. In general, if we are just given a correlation between two variables X and Y, this does not tell us whether it makes sense to regress "X on Y" or "Y on X". In truth only one of these analyses will be appropriate, for only one will yield an equation structure with error term independence—depending on whether in reality it is X that causes Y, or vice versa.

(18)



Here too we can see the causal structure as deriving from the sequence of independent error terms. In this more complex set of equations, 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—this constitutes S as an effect of P and e_S. And in turn 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 constitutes E as an effect of all of P, S and e_E.

I have been using linear regression analysis to illustrate the idea that causal structure might derive from recursive structures of deterministic equations with independent error 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 observable variables $X_1, \ldots X_n$ and error 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

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

Then in general we can view the DASs of these equations as capturing causal structure on the assumption that the error terms are probabilistically independent.

Just as with the linear regression examples I have used, the independence of the error terms is naturally viewed as imbuing the whole system of equations with causal structure. Each dependent variable X has its values fixed by its independent variables and its own error term. The latter adds some X-specific variation to the mean values for X determined by the independent variables, and so constitutes X as an effect of the terms on the right-hand side of its equation.

So I now propose the following reductive analysis of causation:

(25) Causal structures are nothing but recursive structures of deterministic equations with independent error terms.

At the metaphysical level, cause-DASs are simply equation-DASs described in different terms. (From now on I shall read "equation-DASs" as implying independent error terms, in recognition of the fact that their recursive ordering is motivated by this independence.)

In a sense, this analysis of causation combines a *regularity* theory of causal *covariation* with a *statistical* account of causal *direction*. We start with a set of deterministic equations. These specify how certain variables covary deterministically in a lawlike way.[19] 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 error terms be probabilistically independent of each other. This means that the error term in each equation operates independently from the independent variables in fixing values for the dependent variables, and so constitutes the terms on the right-hand side as causes of that dependent variable. (I shall now switch terminology from "error terms" to "exogenous variables"—the requirement is still that these terms appear only on the right-hand sides in a recursive system of deterministic equations, but our interest has now switched from their lying beyond empirical observation to the way in which they metaphysically determine causal structure.)

13. Recovering the Bridge Principles

Let us now go back to the bridge principles that are employed in non-experimental research to infer causal structures from correlational data. The reduction of causation I have just proposed now allows a proper appreciation of their status.

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:

(26) 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.)

Note that this result says nothings 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 error terms be independent.

Still, the result has an obvious causal significance when combined with reduction of causation that I am proposing. If causal-DASs are nothing but equation-DASs, then the Determinism-Independence-Markov Result (26) implies the earlier *Causal* Markov Condition (4)—every variable in a *causal*-DAS 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 Conditions (2, 3) that play so central a role in accounting for the ability of empirical researchers to draw causal conclusions from correlational premises. Two variables in a causal structure will only be correlated if they are casually linked, and this correlation will be screened off if we control for causal intermediaries.

[19] 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.

So the analysis of causation I am recommending inherits the ability of the neo-probabilistic theory to account for the Causal Markov and consequential Linkage and Conditional Linkage Conditions. It is noteworthy, though, that it does not simply *posit* that causal structures will satisfy these conditions, as I did in section 4 above. Rather it *derives* this from a deeper analysis of causation in terms of deterministic equations with independent exogenous variables.

It is also noteworthy that the Faithfulness Condition does *not* follow from the identification of causal structures with equation-DASs. 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. Failures of faithfulness might be an unusual freak, but the kinds of cancelling-out causal structures that give rise to them seem metaphysically perfectly possible.

The reductive analysis I am now defending gets this right. There is nothing in the identification of causal structure with equation-DASs to guarantee that probabilistic independencies should not arise by a cancelling out of parameters. This would be a freakish chance, and to that extent it can be discounted as a serious possibility, but it is not built into the nature of causation. The proposed reduction of causal structures to equation-DASs does guarantee the Causal Markov Condition that tells us correlated variables in a causal structure are always causally linked, but the converse Faithfulness Condition that tells us causally linked variables must always be correlated is only delivered as a reliable rule of thumb. 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.

At this point, it will worth saying something about backgrounds fields. In section 3 above I observed that projectible population correlations will be relative to a background field. The same applies to equation-DASs. The equations and probabilistic independencies that constitute any such 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 *strength* of the correlations between variables in an equation-DASs 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

(17.1) $P = e_P$

(17.2) $S = aP + e_S$

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 tied to P.[20] The linear dependency (17.2) itself, however, can be expected to be more robust than the extent of such variation. There is no preordained 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.[21]

15. Failures of Faithfulness Redux

The Faithfulness Condition might fall into place once we replace the neo-probabilistic theory of causation by a reduction to equation-DASs. But even so failures of faithfulness might still seem to present a challenge to the latter reduction.

My current reductive proposal (25) is that causal structures are nothing but structures of deterministic equations with probabilistically independent exogenous terms. But the possibility of unfaithful cancelling-out independencies ("specious" independencies henceforth) threatens equation-DASs that do not correspond to causal structure.

Consider a possibility raised earlier: high parental income P might actually have a negative influence on examination results E, with this influence being precisely cancelled out by the direct positive influence of well-funded schools S. In that case, we would still have the following underlying equations with independent exogenous terms:

(17.1) $P = e_P$

(17.2) $S = aP + e_S$

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

And this equation-DAS would still determine this causal structure:

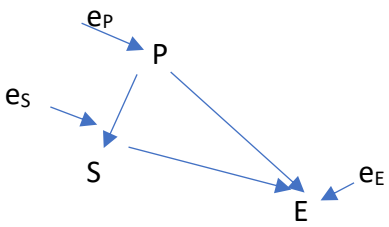
[20] Under the standard assumptions of linear regression analysis, the correlation $r_{P,S}$ between P and S is related to the regression coefficient a by

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

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 (17), for example, faithfulness failure is guaranteed by $a + bc = 0$, whatever the variances.

[21] 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.

(18)



But if we had the hypothesised cancelling-out, with P and E overall unconditionally uncorrelated, then the following pair of equations would *also* be a recursive structure with independent exogenous terms:

(27.1) $P = e_P$

(27.2) $E = e^{**}_E$

The cancelling-out that renders P and E probabilistically independent would mean that the posited exogenous terms e_P and e^{**}_E are also independent—and so, given this equation-DAS, my proposed reduction would thus imply that P and E are causally unlinked—which is not the conclusion we want.

16. Accidental Independencies

A natural first thought in response to this problem is that the probabilistic independency of the "exogenous variables" in equations (27) is an accident. We started off with the genuinely lawlike set of equation (17), and then the happenstantial cancelling out of coefficients a + bc = 0 meant that P and E ended up as probabilistically independent.

So a quick answer to the threat posed in the last section would be to specify that equation-DASs that reductively constitute causation must involve genuinely lawlike exogenous probabilistic independencies, not specious ones that result from an accidental cancelling-out of coefficients, as in the equations (27).

While I think this answer is ultimately sound, it might seem unreasonably quick in the present context of argument. So far I have said nothing about the metaphysical basis of the "genuinely lawlike independencies" in causally respectable equation-DASs, beyond saying that they should have some kind of projectible lawlike status. Given this, it is unclear that I am in any position to dismiss the specious independencies displayed by equations (27) as accidental. After all, they will remain as long as the coefficients in equations (17) cancel out (a + bc = 0). Since I am taking these equations themselves to be lawlike, should I not recognise any independencies dictated by their coefficients as lawlike too?

In fact I shall say a bit more in the next two sections about the source of the probabilistic independencies in causally respectable equation-DASs, suggesting that they have their origin in the decoherence-driven emergence of macroscopic entities from the quantum realm. But there will be no question of developing this thought properly in the context of this paper. So instead of demoting the specious independencies on the somewhat

speculative grounds they lack this kind of quantum backing, let me here offer a different rationale for putting them to one side.

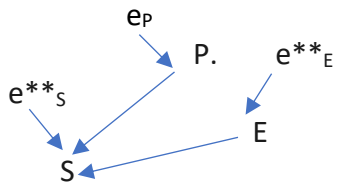
Consider what happens if we try to extend the unwanted specious equation-DAS (27) by also adding a third equation, in line with the mistaken causal structure misleadingly suggested by the specious independencies.

(27.1) $P = e_P$

(27.2) $E = e^{**}_E$

(27.3) $S = fP + gE + e^{**}_S$

(28)



The equation (27.3) now has S as a function of P and E. But the overall system no longer qualifies as an equation-DAS, because the augmented set of equations won't have independent exogenous terms throughout. In particular, the term e^{**}_S won't be independent of e^{**}_E, given that in truth E depends causally on S. (The term e^{**}_S will in effect need to compensate for how E in fact varies in ways that are independent of S.)

Given this, I now propose that we add a further requirement on the kind of equation-DASs that can ground causal structure. They need to be expandable so as to accommodate any further correlated variables in a larger equation-DAS. More precisely, the equation-DASs that ground causal structure must satisfy this condition:

(29) If two variables X and Y in the equation-DAS have their values fixed by probabilistically independent exogenous terms, then for any further variable Z correlated with both X and Y there must be an expanded equation-DAS in which Z appears as dependent on both X and Y.

In our example of specious P-E cancelling-out, there is no way of expanding the original two equations (27.1) and (27.2) into a larger equation-DAS that has S as a function of P and E. If we try to do so by adding equations (27.3), we violate the requirement on equation-DASs that all exogenous terms be probabilistically independent.

Since specious independencies arise specifically when some X causes some Y via two different routes, they will always fall foul of this requirement. Any variable Z that is intermediate on either route will be correlated with both X and Y, yet there will be no equation in which it is a function of X and Y and some further independent exogenous variable. Given that in truth Z will be a cause of Y, the exogenous term in the equation for Z

will need to correct for the ways Y actually varies independently of Z, as illustrated in our P-S-E example.

17. Causal Determinism

The reduction of causation I have proposed 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.

But this appearance is misleading. For one thing, my reduction does not imply that everything is determined, only that *effects* are. For another, 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 might still be the outcomes of chancy quantum processes, and moreover the values of the 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.

Still, even if the deterministic presupposition of my reduction can be rendered consistent with quantum mechanics in this way, that scarcely means that we are compelled to believe it. Why must we suppose that all effects are determined by prior facts, even given that the determination need only come into operation, so to speak, shortly before the time of the effect? What rules out the possibility that some effects are themselves genuinely chancy, with all prior facts leaving a chance of their occurrence different from zero or one?

In my view, the Markov behaviour characteristic of causal structures gives us strong reason to accept that effects are always determined. Recall the Determinism-Independence-Markov result (26) above. This said that *if* the dependent variables are *determined* by other variables with probabilistically independent exogenous terms, *then* any correlations between variables will be screened off by common ancestors or intermediaries. That is, the deterministic nature of equation-DASs with exogenous independence *explains* why causal structures display their distinctive Markov screening-off properties. I take this to argue that causal structures really are deterministic.

This argument is of course not conclusive. Suppose that the totality of prior circumstances didn't determine effects, but just fixed non-zero-or-one chances for them. Why shouldn't that equally allow causal structures to display the Markov property with correlations being screened off by common causes and causal intermediaries? Well, it would *allow* it—but it wouldn't *require* it. And it is striking that in all known real circumstances where a temporally prior state fixes pure chances for two separate results, or where a temporally intermediate state similarly fixes pure chances for some prior and subsequent state, these non-

determining ancestors or intermediaries do *not* screen off the correlations between the states they link.

I am thinking here of the kind of "non-local" quantum correlations displayed by EPR phenomena or the Aharonov-Bohm effect (Healey 1997). Initial intuition leads one to expect that the temporally prior or intermediate quantum states in these cases will screen off the correlations between the states they link. But quantum theory implies the opposite, and experiment confirms this. Surprisingly, the correlations aren't screened off by the linking states. (Indeed, as Bell's inequality shows, the EPR correlations have a structure that means they *can't possibly* be screened off by linking local states, even hidden ones that lay beyond the known quantum states.)

This now strongly reinforces the argument for deterministic causation. Causal structures display the Markov screening-off property. This property is a deductive consequence of deterministic equation-DASs. By contrast, in all known cases where possible screener-offs do not determine the states they link, but only fix non-zero-or-one chances for them, screening off is not displayed. The obvious implication is that deterministic equation-DASs lie behind causal structures. The characteristic Markov screening-off feature of causation is due to the way sequences of probabilistically independent terms *determine* certain results.

(From this perspective, perhaps the strange quantum correlations are less surprising than they first appear. Think of it like this. Deterministic equation-DASs leave no possibility of correlations that aren't screened off by intermediary links. On the other hand, if we don't have this underlying deterministic structure, as in purely chancy quantum mechanical set-ups, then there is probabilistic room, so to speak, for states to become correlated in a way that isn't screened off by intermediary links. And it turns out that, as soon as nature has this room, it uses it to produce correlations that can't be screened off. What's so surprising about that?)

18. The Linkage Conditions Clarified

According to the argument of the last section, non-local quantum correlations are not *causal*. They do not display the characteristic Markov screening-off properties that are necessitated in variables governed by equation-DASs.

This conclusion, however, is in tension with the Linkage Condition as originally stated, which said:

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

If non-local quantum correlations are not causal, then they are immediate counter-examples to this Condition.

Moreover, I argued in section 13 above that the Linkage Condition (along with the associated Conditional Linkage Condition) followed from the Determinism-Independence-Markov Result plus my proposed reduction of causation to equation-DASs. Since the

Determinism-Independence-Markov Result is a straightforward mathematical theorem, this looks bad for my proposed reduction.

It is important, however, that the derivation in section 13 was qualified. It said that all *causal structures* will satisfy the Linkage and Conditional Linkage Conditions, not that all variables whatsoever will. And, in the context of that derivation, causal structures meant sets of variables governed by equation-DASs. Since the variables involved in quantum correlations are not so governed, they are not covered by section 13's derivation of the Linkage and Screening-Off Conditions.

So the argument from section 13 did not in fact deliver the fully general Linkage Condition (2), but rather the qualified:

(30) If two variables *in a causal structure* are correlated, then they must be causally linked.

To see the relevance of this qualification more clearly, suppose that the equations for X and Y in some equation-DAS are

(31) $X = e_x$ and

(32) $Y = e_y$

The requirement that e_x and e_y are probabilistically independent then trivially forces the independence of X and Y. More generally, any two variables in an equation-DAS whose values are fixed by disjoint sets of exogenous variables will similarly have their probabilistic independence forced. That is why my proposed reduction of causal structures delivers the qualified Linkage Condition (30). (And a similar argument will apply to a qualified Conditional Linkage Condition.)

However, there is no reason to suppose the paired variables involved in pure quantum correlations are jointly governed by equation-DASs, so they escape this argument that correlations imply causal linkages.[22]

If the Linkage and the associated Conditional Linkage Condition are not fully general, then where does that leave the non-experimental researchers who rely on them to infer causes from correlations? Should they not worry that their conclusions will be invalidated by

[22] The derivation of the Linkage and Conditional Linkage Conditions from the Causal Markov Condition—every variable in a *causal* structure will be independent of every non-descendant given its parents—itself imposes a restriction to variables in casual structures. When we assumed back in section 4 that nothing is required of causal structures beyond including all common causes of included variables (cf footnote 9 above), then the Causal Markov Condition implied the original Linkage Condition (2) in full generality. This was because any pair of variables that are not causally linked will themselves comprise a minimal causal structure, and the Causal Markov Condition then trivially implies their probabilistic independence. But now we are requiring in addition that the variables in a causal structure be governed by equation-DASs with exogenous independence, and this narrows the scope of the Causal Markov Condition and its consequences so as to exclude quantum correlations.

exceptions to these principles? However, we have seen no reason to doubt that these principles apply to all macroscopic variables of the kind at issue in typical non-experimental research. As we saw in section 9 above, none of the non-quantum counter-examples to the Linkage or Conditional Linkage Conditions is compelling.

It is attractive to suppose that governance by equation-DASs is built into the emergence of macroscopic entities from the quantum realm. As we have seen, quantum phenomena do not generally display the Markov behaviour dictated by equation-DASs. But once we have the effective decoherence occasioned by macroscopic interactions, this seems to bring with it the factorisation into probabilistically independent influences captured by equation-DASs, and therewith the satisfaction of the Linkage or Conditional Linkage Conditions. (Note once more that this picture does not require the macroscopic realm to be fully deterministic, only that macroscopic events be determined by definite factors by the time they occur.)

Still, this is not the place to pursue these quantum speculations. For present purposes we can rest with the observation that we have no reason to doubt that the Linkage and Conditional Linkage Conditions apply to all macroscopic variables.

19. Unfinished Business

Much philosophical work on causation over the past two decades has appealed to deterministic "causal models" of particular situations to analyse token claims about *actual causation*[23] and *counterfactuals*[24]. These models posit directed deterministic relationships, standardly portrayed by arrows, between actual and possible values of variables displayed by particular situations. The aim is to formulate recipes that will allow us to read off from the models which events some given result was actually caused by or counterfactually dependent on.

The analysis of this paper complements this work. While much progress has been made on the way causal models can help analyse actual causation and counterfactuals, the directed relationships they invoke have largely be taken as given. These relationships are represented as arrows, and are often called "*structural* equations", but there is no agreed explanation of the arrows and in particular for reading the co-variation of variables as asymmetrical (cf Beebee and Menzies 2020 section 5.2). The present paper fills this lacuna. It shows that we can view the structural equations in the models as representing deterministic equations over variables displayed by given types of entities in background fields, with the direction deriving from the probabilistic independence of the exogenous variables in these equations.

This account of the meaning of the arrows in causal models raises further questions. Actual causation and counterfactual dependence are relationships between pairs of token events. I am now suggesting that the causal models be read as specifying general relationships between *types* of events *relative* to a given background field. This leaves us with choices about which models are "apt" for a given pair of particular events. What class of cases do we want to consider our token events instances of? The choice of a background field will in

[23] Hitchcock 2001, Halpern 2016, Blanchard and Schaffer 2017, Weslake forthcoming
[24] Galles and Pearl 1998, Schulz 2011, Briggs 2012

effect determine which possible variations away from actuality are covered by the model. Someone contracts a viral infection and dies. What alternatives, if any, to their actual medical treatment, or actual immune response, or actual genetic make-up . . . should be entertained? This paper leaves open how far such choices might be objectively constrained, and how far conclusions about actual causation and counterfactual dependence might be sensitive to them.

A rather different set of questions relate to the generalizations that issue from the kind of correlational studies we started with. “Smoking causes cancer.” “Parental income affects examination results.” As I said at the beginning of the paper, the notion of *causal influence* in play in such claims needs unpacking. My analysis now opens the way to analysing such claims as generalizations over facts of directed minimal sufficient conditionship. (Cf Weslake forthcoming.) For example, “smoking causes cancer” can be read as saying that smoking level is *sometimes* an essential part of a set of causally upstream variable values that together determine cancer.

Correlational studies will also typically deliver probabilistic conclusions about the *importance* of such generic causes. For example, we might be told *how much* the probability of cancer is increased by different levels of smoking in some background field. From my perspective, this information then tells us how probable it is that a given level of smoking will be decisive to the presence of a directed minimal sufficient condition for smoking. Information of this form has obvious relevance to decisions—for example, a decision on whether it is worth quitting smoking to avoid cancer.

Some philosophers hold that causation cannot be analysed without appealing to the concept of human *action* or some prior category of *intervention*. This is a topic that deserves fuller discussion, but the analysis of this paper argues that this claim is a mistake. 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 or intervention were prior to causation. After all, humans and their activities are part of the causal world, not prior to it.

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. 1972 *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

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

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

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

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

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

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

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

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

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

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

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

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

Schulz, K., 2011 “‘If You’d Wiggled A, then B Would’ve Changed’: Causality and Counterfactual Conditionals” *Synthese* 179: 239–51

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

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

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

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*

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