**Explaining Causation**

**David Papineau**

**Bristol Department of Philosophy**

**May 18 2023**

***Causal Inference***

Despite what we are told, correlations do often imply causation (or close enough).

Such implications are the business of *“causal inference”*—a fast expanding area of applied science.

None of the existing metaphysical accounts of causation (counterfactual, regularity, powers, process, . . .) do anything to explain why causal inference techniques work. I will offer an account that will do this (and explain many other features of causation as well).

Take this case. Kidney disease (K) and alcohol drinking (A) are both correlated with hypertension (H), but neither is correlated with the other. The only causal structure over these three quantities consistent with this is:

 K A

 H

(If K caused A or A caused K, either directly or through H, then we would expect to find a correlation between K and A. And similarly, if H caused both K and A, then we would again expect to find a correlation between K and A.)

***Bayesian Network Principles***

We need *Bridge Principles* linking correlations with causes—"Bayesian Network” principles.

Defn: quantities X and Y are *causally linked* if X causes Y (possibly indirectly via a path involving intermediaries), or Y causes X (again possibly indirectly), or X and Y have a (possibly indirect) common cause—but not if X and Y only have a common effect.

One bridge principle—Reichenbach’s Principle—takes us *from* correlations *to* causal links.

If two quantities X and Y are correlated, then they must be causally linked by one or more causal paths.

Moreover, if we control for at least one element Z in each causal path linking X and Y, then X and Y will become conditionally independent: X ⊥ Y | {Z}.

The other bridge principle—the Faithfulness Condition—takes us from *absence* of correlation to *absence* of causal links.

If two quantities X and Y are uncorrelated, then they are not causally linked.

Moreover, if two correlated variables X and Y become conditionally uncorrelated when we control for other variables {Z}, then they are not causally linked by any chains of variables that do not contain any of {Z}.

These two principles suffice to fix causal order given sufficiently rich sets of correlations.

Not all sets of correlations are rich enough to fix causal order. For example, A corr B, A corr C, B corr C, A indep C | B, are consistent with all of:

A 🡪 B 🡪 C; A 🡨 B 🡨 C; A 🡨 B 🡪 C – though not with A 🡪 B 🡨 C.

Still, if we add that some D is corr with A, B, C with the D-A and D-C correlations screened off by B, then the only causal possibility left is:

 D

 B

 A C

There’s a general theorem: if some set of correlations fails to fix causal order, there is always a richer possible set that will.

***Confounding Common Causes***

A confession. I’ve been cheating a bit. The Bridge Principles only fix causal order given rich enough correlations *if the quantities we’re dealing with are causally complete*—in the sense that they don’t omit common causes of the quantities they include.

(More technically, the general theorem above doesn’t say my Bridge Principles will fix causal order given causal completeness. It says the Markov and Faithfulness Conditions will do so. But the Markov Condition follows from the conjunction of Reichenbach’s Principle and causal completeness.)

This is because our inferences above assumed that if A and B are correlated, and this correlation isn’t screened off by any other quantities *in our analysis*, then A must be causing B or vice versa. But of course this assumption will be false if A and B are correlated due to a common cause. (Maybe Kidney disease and Hypertension are joint effects of Obesity.)

Still, the worry goes away if our variables are causally complete. Then it is indeed true that that if A and B are correlated, and this correlation isn’t screened off by any other quantities in our analysis, then A must be causing B or vice versa (albeit perhaps less directly that our analysis displays).

Now, this is a big practical issue in the causal inference world. Can we ever be confident that a study is casually complete? Different disciplines have different attitudes to this danger. Roughly speaking, epidemiologists and social scientists hold that the danger of hidden confounders can often be surmounted by a judicious selection of surveyed quantities. Medical scientists and economists, by contrast, tend to feel that surveys can never eliminate the danger, and that it needs to be forcibly addressed by conducting “randomized controlled trials”.

But we can by-pass this practical issue here. We have a more pressing philosophical question. What in the nature of causation allows correlations to fix causation given causal completeness? (Indeed what in the nature of causation allows the results of “randomized controlled trials” to fix causation?)

***Probabilistic Theories of Causation***

Probabilistic theories of causation aim to reduce causation directly to correlational patterns. A causes B iff this follows via Bridge Principles from the correlations.

(What about the requirement of *causal* completeness? Say rather: . . . iff this follows via Bridge Principles from some correlations *and* this isn’t overturned by any richer set of correlations involving more variables.)

Hausmann: A causes B iff they are correlated and everything correlated with A is correlated with B and something correlated with B isn’t correlated with A.

(This rules out A and B having a common cause because then something correlated with A wouldn’t be correlated with B.)

Problems for probabilistic theories:

1. Single-cases. Smoking causes cancer. Jim smokes and gets cancer. But did his smoking cause his cnacer?

2. Failures of Faithfulness. Hesslow’s example: Birth control pills (B) cause Thromboses (T), but prevent Pregnancies (P) that also cause Thromboses. If the two processes cancel out, we can get a zero B-T correlation, which will wrongly imply B 🡪 P 🡨 T.

3. Failures of Reichenbach’s Principle. Venice water and London bread. EPR correlations.

All these problems argue that real causation lies deeper than the correlations, which are in truth only symptoms of the underlying fine-grained causal structure.

***Structural Laws***

Older statisticians took the correlations to be evidence for the underlying “structural equations”.

X 🡨 ex

Y 🡨 aX + ey

Z 🡨 bX + cY + ez

If these are “equations”, can’t they be reordered so as to give us different dependencies? But then we’ll lose a crucial feature—the exogenous variables will no longer be probabilistically independent of each other.

This independence is crucial to the use of these equations. It’s why we can use X to predict Y while averaging over the error terms, and so on.

I say this captures what it is for X to cause Y, and X and Y in turn to cause Z.

We don’t need real-valued quantities or linearity. We can generalise to any set of recursive deterministic laws Xi 🡨 F(X1, . . . Xi-1, Ei) over dependent quantities X1, . . . Xn and exogenous quantities E1, . . . En:

 Xi is a *type-cause* of Xj if and only if it is an ancestor of Xj in a recursive structure of deterministic laws with independent exogenous terms.

***Recovering the Bayesian Network Principles***

The above analysis implies Reichenbach’s Principle for quantities governed by RLIs (Recursive structure of deterministic Laws with Independent exogenous terms). To be correlated, they can’t descend from disjoint set of exogenous quantities.

But it doesn’t imply the Faithfulness Condition. While probabilistic independence will normally be due to descent from disjoint set of exogenous quantities, it can also sometimes result from freaky cancelling-out.

So this gives us a deeper understanding of Bayesian Network causal inferences. We’re using the correlations to tell us something about structures of underlying laws.

Note that Reichenbach’s Principle is only guaranteed for quantities governed by RLIs. EPR-correlated quantities aren’t so governed (because their equations aren’t deterministic) and Venice water and London bread levels aren’t so governed (because their equations don’t generalise over kosher objects).

NB recovering the Bayesian Network principles only requires Causal Structures 🡪 RLIs. It’s a further claim that RLIs 🡪 Causal Structures. But this reduction is attractive. Not clear anything it leaves unexplained.

In effect, we’ll then reduce causation to laws and probabilities. Plug in your own favoured accounts of these.

Nothing about temporal ordering has been assumed in this analysis so far. But in fact causes will always precede effects. It would be nice to have an account of this temporal asymmetry that unites it with the temporal asymmetry of thermodynamics. (Sean Carroll has a talk that attempts this.)

***Single Cases***

These laws (RLIs) are by their nature generic. How do they relate to *single-case* causation and counterfactual dependence? Here I appeal to the wealth of work in the Lewisian tradition that has devised recipes for reading answers to these questions off from “causal models” that specify how dependent variables are deterministic functions of others.

For example, Susy and Billy and the Smashed bottle:

ST SH

BT BH S

What do the arrows mean here? Existing work takes them to portray either some primitive causal dependence or complex counterfactuals. I can do better: the models answer to how the variables are related in an RLI.

This opens the way to explaining counterfactuals (and actual causation) in terms of RLIs. *Interventionist semantics* for counterfactuals: set the relevant variable to the counterfactual value, leave all non-descendants unchanged, crank the equations . . . The bottle Shattering doesn’t depend counterfactually on Susy Throwing.

For actual causation, the story is trickier (the Shattering was actually caused by Susy Throwing). Brad Weslake has suggested that this should hinge on whether the structural laws and the actual facts mean that C is an ancestor of E via the relation of causal INUS conditionship. (Given Billy didn’t Hit, Susy Hitting was an INUS condition of Shattering; and her Throwing was an INUS condition of her hitting.)

***Smoking Causes Cancer***

Let us go back to the kind of claims we might infer from correlations via the Bayesian Network principle—eg “Smoking causes Cancer”.

I have shown that we can regard this as telling us that Smoking is one of the type-ancestors of Cancer in the recursive structure of directed laws relating them (RLI).

Q: why are we so interested in such coarse-grained generic facts? As we’ve seen, they are likely to leave us in the dark about single-case realtions.

A: because coarse-grained generic facts like “Smoking causes Cancer” are just what we need to know when deciding on action.

We can take “Smoking causes Cancer” to tell us that Smoking *sometimes* creates a sufficient causal condition for Cancer when one wouldn’t otherwise exist—and thus that it’s not a good idea to Smoke if you want to avoid Cancer.

What’s more, Bayesian Network inferences can also deliver numbers which tell us *how likely* it is that Smoking will create such a sufficient condition for Cancer, and thus how much reason you have to not Smoke to avoid Cancer.

NB1. It is not the *gross correlation* between S and C – Pr(C|S) vs Pr(C)—that gives us this information, but the correlation *conditional on all the other causes* O of cancer--Pr(C|S&O) vs Pr(C|O). If S is itself correlated with O, then we’ll see a gross correlation even if S doesn’t itself ever create a sufficient causal condition for Cancer.

NB2. What we want to guide action is how often S creates a condition on which C *counterfactually depends*, not a condition which *actually causes* C. (If all Susy cares about is the bottle Shattering, she has no reason to Throw, even though she’ll actually cause the shattering if she does.)

NB3. We can show that, if Pr(C|S&O) > Pr(C|O), then it C must sometimes counterfactually depend on S. But it’s not true that, if Pr(C|S&O) = Pr(C|O), then C never counterfactually depends on S. That’s because of Faithfulness failures. There might be multiple routes between S and C, and the conditional correlation only shows us the net effect of them all. Still, despite this, the conditional correlation will nevertheless be a good guide to action. (If all you care about is avoiding Thromboses, you’ve no reason to take/avoid Birth control pills in Hesslow’s set-up.)

***Agency Theories***

I am explaining rational action in term of the likelihood that the desired result will depend causally-counterfactually on the action. (Note how this explains rational action in terms of counterfactuals in terms of causation in terms of RLIs . . . I’ll say “causal-counterfactually” to remind us I’m analysing counterfactuals in terms of directed type-causation.)

*Agency theorists* will say this is all the wrong way round. Won’t we do better to explain causation and its asymmetry in terms of *when it’s rational to act*?

Let me digress on *“Interventionism*.” C type-causes E iff, *were* C to be produced by an intervention, there would be a correlation between C and E – where an “intervention” is some way of producing C that render C probabilistically independent of the *other causes* O of E.

This is true enough, I say.

It follows from the points I have established so far, including my analysis of counterfactuals (putting to one side the problem that Woodward’s analysis fails in the “only if” direction because of Faithfulness failures, an issue that he struggles to deal for the first half of his book).

But it’s obviously not much of an analysis of causation, given it has to invoke causes in the definition of “intervention”.

Nor is it, despite appearances, an analysis of causation in terms of agency.

Pearl says that gross correlations – Pr(C|S) – are fine if you want to predict what will follow what you observe – will an observed Smoker get Cancer? But he says you need a different number – Pr(C|S&O) – if you want to “predict the result of an action”. Perhaps there is a gene O responsible for both S and C, in which case the S-C association is *spurious*, and you don’t need to quit Smoking.

What Pearl says makes little sense. What he means is nothing to do with prediction, but rather that the latter number Pr(C|S&O) is what you want for *decision* – to work out whether it’s *worth* quitting Smoking to avoid Cancer.

In effect, Pearl is just advocating Causal Decision Theory over Evidential Decision Theory. He’s not explaining causation in terms of action, just saying you should choose according to the probability your action will *cause* the result, not according to the gross probabilistic association of the result with your action.

Serious agency theorists (Price, Fernandes, Glymour, Hitchcock, . . .) want to say causation is the relation C bears to E *when it’s rational to do C in pursuit of E*, and then explain the latter *without* bringing in causation—rather, they explain it in terms of *C rendering E more likely*. But then they have their work cut out to not advocate acting on spurious symptoms—and it’s not clear they can.

Another way of putting the point. If human actions were generally *interventions*, in the technical sense, then we’d have a chance of analysing causation in terms of them, as above. But they aren’t—factors that independently cause E often influence choices of C as well, which will produce spurious C-E correlations, and the agency theorists can’t any more say causation is the relation C bears to E whenever acting to produce C renders E more likely.