The Seductions of Interventionism

David Papineau
Birmingham 21 January 2020

1. Epidemiologists, econometricians, and other non-experimental scientists have long used correlations patterns to infer causal conclusions. It has always struck me as a scandal that metaphysical theories of causation cast no light on why these techniques work. Now the work of Judea Pearl, assisted by Jim Woodward, has brought these techniques to the attention of philosophers. Yet the philosophers still avoid the basic issue of why they work.

I shall (i) explain the basic relations between correlations and causation (ii) show that resistance to reducing causes to correlations is misplaced (iii) expose the confusions engendered by the notion of an"intervention".

2. Standard multivariate linear regression:

$x = e_x$
$y = ax + e_y$
$z = bx + cy + e_z$

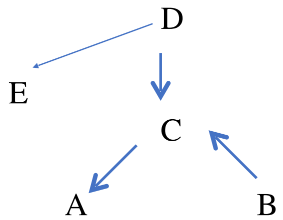
Intuitively, these equations imply that x causes y and that x and y cause z.

How so? Can't we rewrite so y depends on z and x, say?

$z = (b + ca)e_x + ce_y + e_z$
$x = -z/(b + ca) + (cae_y + e_z)/(b + ca)$
$y = z/c – bx/c – e_z/c$

Yes but now the error terms aren't *probabilistically independent*. The underlying intuitive idea is that *structural "equations"* must satisfy the constraint that the exogeneous variables are independent—and then they will represent causal structure.

3. Pearl shows that *if* the exogeneous variables in any recursive system of equations are independent, *then* all the variables will satisfy the *Markov Condition*: every variable in the DAG (directed acyclic graph) will be independent of all other variables except its descendants conditional on its parents.



4. So far Pearl's theorem doesn't say anything about *causation*. Just a graphic way of representing probability distributions. But it's natural to take the arrows here to represent causation. This requires assumptions to link correlations and causation.

Causal Markov condition: in a *causal* structure, every variable will be independent of all other variables except its descendants conditional on its parents. (Tells us *correlated variables are causally connected*, plus some stuff about screening off (where "causally connected" means one causes the other or they have a common cause).)

Causal Faithfulness: uncorrelated variables are *causally* unconnected.

5. These are what practising non-experimental scientists use to infer causation. *Worry*: the scientists don't normally look at a full set of variables giving deterministic equations. *Answer*: this doesn't matter; as long as you observe enough variables to have to have exogenous ones independent, and don't leave out any common causes, your conclusions will be robust w.r.t. bringing in more fine-grained variables. *Worry*: Any analysis will assume a "causal field" in which it's taken for granted that background variables have fixed values. *Answer*: We can always broaden our analysis and look at the causes of these variables' values too.

6. So *why* do these techniques work? Why accounts for causation manifesting its structure in correlations? Pearl says it is "a gift from the gods". Philosophers say nothing.

The obvious answer is that this is what causal structure *is*. It's precisely the probabilistic independencies among exogenous variables that adds directional causal structure to lawlike connections among variables. (Give me another explanation.)

For an explicit reduction: If A is correlated with B, then A is a cause of B iff everything correlated with A is correlated with B, and something correlated with B isn't correlated with A (cf Hausman, *Causal Asymmetry*).

7. Pearl is adamantly against this reduction, and nearly all philosophers follow him. One relatively superficial reason for Pearl's resistance is that he's from a tradition of statisticians/computer scientists who want to read the arrows in the DAGs as simply representing probability distributions and avoid any dodgy causal notions. Pearl rightly responds that causal structure is important and must be attended to. Still, we can *reduce* A to B even if A *conceptually* transcends B. Reduction doesn't mean elimination.

8. And then there's "no causes in, no causes out". The correlational structure underdetermines the causal structure. Widely believed but not true. Sure, some sets of correlations underdetermine causal structure. And in such cases practical researchers use prior knowledge, time order, etc to help figure out causal structure. But in principle wider sets of variables can always resolve the underdetermination.

Suppose we have A corr B, A corr C, B corr C, A ^ B/C.

('A ^ B' means A and B are uncorrelated; 'A ^ B/C' means A and B are uncorrelated *given* C, ie C 'screens off' the correlation between A and B.)

This might indicate (i):

$$\begin{tikzcd} & C \arrow[dl] \arrow[dr] & \\ A & & B \end{tikzcd}$$

But it could also mean (ii): $A \rightarrow C \rightarrow B$

or (iii): $A \leftarrow C \leftarrow B$

(but *not* (iv):

$$\begin{tikzcd} & C & \\ A \arrow[ur] & & B) \arrow[ul] \end{tikzcd}$$

Still, suppose we are now also given D ^ B, D corr C, D corr A, D ^ A/C. Then this indicates, as the only option, (v):

$$\begin{tikzcd} & D \arrow[d] & \\ & C \arrow[dl] & \\ A & & B \arrow[ul] \end{tikzcd}$$

There's a general theorem that for any underdetermined possible structure there's a possible wider embedding structure that will determine the causal order in the original structure. Yet Hitchcock in his SEP article on "Causal Models" omits to mention this, and instead merely says: "If we don't have information about time ordering, or other substantive assumptions restricting the possible causal structures . . ., then it will not always be possible to identify the causal structure from probability alone." (Cf Hausman's definition above; it presupposes that there will always *be* independent sources of variation to distinguish effects from causes. Seems reasonable to me.)

9. Issues that need addressing, but not here:

(a) Counterexamples to Causal Markov Condition: London bread prices and Venice water levels.
(b) Counterexamples to Causal Faithfulness: freaky cancelling out producing misleading independencies.
(c) Actual Causation: can't read actual causation off from general DAG facts and actual variable values, because of pre-emption, etc.
(d) Laws. The reductive project assumes underlying deterministic laws and also correlational laws. Is there an account of these laws that fits the project but does not pre-empt it?
(e) Does quantum indeterminism make a different to the reductive project?

10. What's all this to do with *intervention*ism? Well, enthusiasts for Pearl's work (including himself and Woodward) say a number of (connected) (and bad) things:

- You need different computations to *predict* what will happen if you intervene on C as opposed to just observing C. (Hitchcock: "Often, however, we are interested in predicting the value of *Y* that will result if we *intervene* to set the value of *X* equal to some particular value *x*" SEP "Causal Models".)

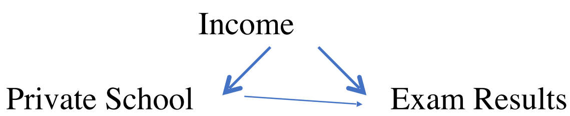
- Causal C-E connections are those which are *robust* with respect to attempts to use C to influence E. (Woodward: '. . . the [spurious] correlation between I and L will not be stable or invariant under efforts to use I to control L' MTH 32.)
- We can *analyse* causal as opposed to other correlations as those which it is objectively rational to act on. (Woodward: 'the distinction between information about correlations and information about relationships that will support manipulations . . .' MTH 34.)

11. It is of course true that causal connections are good to act on and spurious correlations are not. But this is a fact about rational action that is metaphysically posterior to the nature of causation. Rational action means *acting on causes* (in line with causal decision theory). This fact thus adds nothing metaphysical to anything said so far.

12. The notion of an "intervention" is a source of much confusion. When you are deciding whether to do C in pursuit of E, you shouldn't consider the gross correlation between C and E, but the *correlation conditional on the other causes of E*. (The coefficient "b" in our original set of regression equations, was a measure of the association of z with x *conditional on* y.) And that is the correlation you would have if C were what's technically defined as an *"intervention"*, that is, if it were decorrelated from the other causes of E.

Still, all this is perfectly in accord with the above metaphysics of causation plus causal decision theory. These together say act on causal correlations, that is, C-E correlations conditional on the other causes of E.

13. To avoid confusion, it's crucial to appreciate that *actions* are not "interventions"—it's quite normal for them to be correlated with E's other causes.



Because of this, it's *not* true that

- You need different computations to *predict* what will happen if you *act* on C as opposed to just observing C.
- Causal C-E connections are those which are robust with respect to *actions* that use C to influence E.
- We can analyse causal as opposed to other correlations as those which it is objectively rational to *act* on.

14. It would be different if we could show that actions really are interventions (within some reference class). This is in effect what tickle-style defences of evidential decision theory aim to show. But Woodward doesn't even try to do this. When push comes to shove, he admits "intervention" is a different technical notion from action. ("There's the bit where you say it, and the bit where you take it back.") And this then means that all the talk of "interventions" adds nothing to the correlation-based metaphysics of causation.