

Reply to Steele and Pearl

For *Philosophy and Economics*

Nancy Cartwright

Both symposiasts understandably jump right into their favourite topics, modularity and Pearl's very own do-calculus. But these involve a very small part of the lessons the book hopes to draw, so I will describe the basics of the book before addressing their comments.

Pearl it seems wants my book to do what his does, and more. But mine has a very different programme, which he ignores, and moreover I argue that what Pearl wants can't be obtained. *Hunting Causes and Using Them: Studies in Philosophy and Economics (HC&UT)* has two announced aims. First a defence of causal pluralism. There are a variety of different kinds of causal systems; methods for discovering causes differ across the different kinds of systems as do the inferences that can be made from causal knowledge once discovered.

If causal pluralism is right, Pearl's demand to tell economists how they *ought* to think about causation is misplaced; and his own are not *the* methods to use. They work for the special kinds of systems whose causal laws can be represented as he represents them. *HC&UT* argues these are not the only kinds there are, nor uncontroversially the most typical.

My second major thesis is, as Steele records, that metaphysics, methods and use must walk hand-in-hand. The methods used to infer causal claims must underwrite the uses to which those claims are put. Metaphysics can glue the two together. We infer the charge of a particle by measuring its deflection in an electromagnetic field. Then we use the measured charge to predict how strongly the particle will attract and repel others. Electromagnetic theory, which tells us 'what charge is', justifies the leap between the two.

What about causation? At the end of the studies of accounts of causation collected in *HC&UT* I arrived at a frightening conclusion: For causality, we don't have any glue. Nothing available, either empirical or philosophical, shows how our methods for inferring causes justify the uses to which we typically put causal knowledge. Theories in both philosophy and economics tend to be either too close to method – e.g., the probabilistic theory of causality and related causal-Bayes-nets, invariance accounts, and even Heckman and Pearl counterfactuals – or too close to use – as with David Hendry and Kevin Hoover.

The book has three parts. Part I defends my two central theses. It includes a paper on warranting causal claims that makes a distinction that is catching on in evidence-based policy discussions between methods (like RCTs, Bayes-nets, some econometric modelling and deduction from theory) that clinch their results – if the assumptions of the method are met, the results deductively imply the tested hypothesis – versus those that merely vouch for their results (like qualitative comparative analysis, case-control studies and ethnographic methods).

Part II focuses on the two accounts most vocally defended as providing universal characterizing features of causality: Bayes-nets and invariance/modularity methods. Contrary to Pearl there is but one paper dedicated to Bayes-nets and again contrary to Pearl the book makes clear what is meant. In keeping with my focus on hunting causes, I discuss Bayes-nets methods for causal discovery and in keeping with my interest in clinching methods, I focus on methods that are *provably valid* for inferring causal relations taking as axioms the three standard causal-Bayes-nets assumptions: minimality, causal Markov (CMC) and faithfulness. Chapter II.2 of *HC&UT* is titled 'What is wrong with Bayes nets?' Answer: nothing. What is wrong is taking them to be universally applicable. They apply where their axioms hold and, I argue, each of the axioms fails in a variety of real cases.¹

¹ E.g. CMC may readily fail if there is selection bias. This is a particularly interesting case because defining selection bias in a usable way that does not presuppose knowledge of just those parts of the causal structure under investigation is tricky, perhaps even impossible.

Besides, Bayes nets have powerful virtues. The axioms provide a ‘metaphysics’², or ‘implicit definition’ for the causal relations under study.³ Various methods for causal discovery are then provably valid whenever the axioms are satisfied. So metaphysics and method are properly joined.

And use? Steele sees this as part of the interpretation of the arrow in a causal DAG. But the axioms already constrain its interpretation. So it is necessary to show that the axioms imply the intended interpretation or at least that the two are consistent.⁴ I myself take the axioms to provide all the interpretation there is and the uses to be exactly those that can be proven, thus marrying metaphysics, method and use.⁵ So, consider a DAG and a probability measure for a situation satisfying the axioms. Suppose the causal structure and probability change. The axioms provide a theorem machine to predict what follows. Modularity theorists focus narrowly on one special kind of change. But the axioms can generate results for all sorts of changes. We thus have a powerful tool for prediction. Powerful, but epistemically demanding: The predictions are only as secure as the DAG-cum-probability, our model of the changes and the assumption that the axioms are satisfied.

Aside: note that I don’t claim that causal-Bayes-nets suppose indeterminism. To the contrary, the independence of the ‘error’ terms, which is necessary for CMC, is most natural assuming determinism, supposing ‘error’ terms represent omitted causes. Suppes offered a probabilistic theory of causality, not a theory of probabilistic causality. My point in comparing the two is that faithfulness and CMC were already heavily criticized in Suppes’s theory, the first, from Simpson-paradox examples, the second because it is not appropriate if causes *are* probabilistic.⁶ So, like Suppes

² Steele would call this a ‘semantics’. Note that, from my pluralist point of view, this is a metaphysics for the particular causal relation under study, not for *the* causal relation wheresoever it appears.

³ This is much like the way the axioms of a physics theory implicitly define the theoretical concepts involved in them. In the case of causality and Bayes nets, the antecedently understood concepts are primarily probability concepts.

⁴ In the latter case one could either maintain that the connection is supported empirically or insist that only cases where the further manipulation conditions obtain are entitled to be called causal. I take it neither Steele nor I favour this last claim.

⁵ With respect to James Woodward’s interpretation that Steele mentions, *HC&UT* does a similar thing: It provides a ‘representation theorem’ showing that Woodward’s ‘level invariance’ interpretation is appropriate to causal laws systems where causation is asymmetric, irreflexive and all true functional relations derive from the basic causal laws.

⁶ Which Suppes would surely want to allow since for him probabilistic/deterministic are distinctions relative to a model.

theory, causal-Bayes-nets don't hold for many causal relations. (Readers of *Economics and Philosophy* may be interested to note that my criticisms of faithfulness are in line with those of Kevin Hoover.)

Part II also contains two chapters showing that Daniel Hausman and James Woodward's two separate attempts to derive CMC from modularity fail. This would have been a nice argument for their anti-pluralist claims that modularity is *the* key to causality. Despite lengthy and determined attempts, though, their derivations are invalid.

Part III focuses on economics – which readers of *Economics and Philosophy* might be most interested in. It studies accounts by Hausman, Hendry, Heckman, Hoover,⁷ T.F. Cooley and Stephan LeRoy, Julian Reiss, and Herbert Simon, and includes a chapter on using economic models to learn about causal relations in the world. This latter depends on my long-standing claim, contrary e.g. to Robert Sugden, that economic models can be used to discover the contributions of causal capacities.⁸

Turn now to Steele's and Pearl's central topic, modularity. Prima facie modular systems seem very special. But a number of authors suppose that modularity is the hallmark of causality. Besides Pearl and Woodward, these include Hoover, possibly Simon, Cooley and LeRoy, and Hausman. I have two objections to the usual claims about modularity.

First, it is not a hallmark of causality. Recall the Philips curve, a canonical example of a non-modular causal connection – one that, ala Robert Lucas, breaks down under attempts to manipulate the cause (inflation) to control the effect (unemployment).

Consider a Phillips curve:

$$* \quad y_t \text{ c= } \theta\beta[p_t - p_{t-1}] - \theta\beta\pi + y_{pt},$$

⁷ The chapter on Hoover, reading directly from his own definitions, attributes to him a remarkable and original use-based theory of causality but one that is concomitantly not so responsive to standard methods for ascertaining causes. Hoover himself claims this is a mistaken understanding of what he intends. This is a topic for extended discussion, but not in this symposium.

⁸ Assuming independent evidence that the causes have a stable contribution to make in the first place.

where y_t is output at t and p_t is price at t , so $[p_t - p_{t-1}]$ is a measure of inflation. According to * increases in inflation produce increases in output. Supposing that increases in output produce increases in employment, * describes a trade-off between inflation and unemployment. But it is of no use for policy says Lucas: How much output suppliers produce is determined by the price they expect their good to sell for and what they expect their expenses to be. In the Lucas model, average price for goods in the economy is a proxy for expense. The amount of a good supplied is then caused by the ratio of the price of the good to the expected economy-wide price for goods. Lucas assumes suppliers are good guessers about the economy-wide price: The price they expect is the average that obtains. So output of a good is determined by the ratio of the price of the good to the mean of economy-wide prices. Price increases thus cause increases in output which in turn cause decreases in unemployment.

What happens if the government manipulates inflation? Assuming *, unemployment goes down. Not so, Lucas argues. A rise in price for a product in the numerator prompts an increase in output only if it is not offset by the increase in average prices in the denominator – a fact that in * is concealed in θ . If suppliers predict the average price, the denominator goes up too; indeed, if it goes up faster than the numerator, the intervention can even increase unemployment. As Lucas tells it, modularity fails.

There are two standard responses. First: shaky equations just aren't causal. To reply I turn to examples like the toaster and the carburetor where other conventional criteria – pushes, pulls, energy interchanges – argue that the connections are causal. The lemonade-biscuit machine is an example where judgments about causality based on mechanical criteria go opposite to those from a manipulationist view. My verdict is pluralism: Systems can be causal in different ways.

The second response is Steele's: Modularity is more common than I think. I agree I had no business saying modularity *generally* fails, especially since I think the idea of a default position is mistaken. In cases where it matters, do your best to figure out what whether modularity obtains.

Still, Steele must have more luck with machines than I. I destroyed my toaster jiggling bits inside. And my car mechanic always breaks one thing when he fixes another. Nor

am I surprised. We want the causal processes in our day-to-day machines to be stable across reasonably hard use. The very shields that protect them make it hard to manipulate the internal causes separately. Also, typical machines, as well as many biological systems, are like the Lucas example: The causal connections under study depend on the stability of an underlying generating structure.⁹ As with the carburetor, with efficiency in mind, we design structures to guarantee a number of causal processes at once – but doing so makes it difficult to change one by itself.

The most important question about modularity, however, is: Why want it? Ease of repair, as Steele argues, is one good reason and I think his work here is important both practically and philosophically. Another is that modularity makes causal claims testable. That's why I say 'epistemically convenient' rather than 'modular'. Concomitant variation is well-known to be a weak indicator for causality. Given epistemic convenience it can become a sure test. This is the main point of the chapter on modularity.

I study *epistemically convenient linear deterministic systems (ECLDSs)*, which are much like those Pearl studies. Modularity is secured by special variation-free causes for each effect that cause nothing else in the system except by causing it. *HC&UT* shows that for an ECLDSs, regression equations that give correct predictions for the effects as these special causes change one-by-one¹⁰ are *causally correct* – a great feature to have since we have good tools for estimating regressions.¹¹ There is also a weaker result.¹² ECLDSs pair each effect with a cause all its own. Then functionally correct equations where none of these special causes appears where it does not belong are causally correct.

It is a shame that Pearl did not notice these results since they add crucially to his programme. The conditions defining ECLDSs provide the 'metaphysics' of the causal

⁹ I call this 'underlying structure' a 'nomological machine' (cf. my *Nature's Capacities and their Measurement*, 1989, OUP) because it gives rise to law-like regularities. Steele and others sometimes call it a 'mechanism'. I avoid 'mechanism' because it has too many different meanings.

¹⁰ As they can since they are variation free.

¹¹ What is it for a regression equation to be causally correct and what is it to pass the test for concomitant variation? I think answering either of these, as I do in *HC&UT*, makes a small contribution to getting our causal methodology straight.

¹² In the chapter 'Getting Causes from Probabilities: Cartwright on Simon on Causation'.

relations Pearl studies; the do-calculus, a machine for predicting, hence use. But method? How do we first discover the causal system? Simple econometrics teaches how to identify ECLDSs; my two theorems provide sufficient conditions for an identified system to be causally correct. When these conditions are satisfied, we know we are entitled to the predictions the do-calculus provides.

Be careful though. Chapter 9 of *HC&UT* is all about counterfactuals in economics; it discusses Heckman, Cooley and LeRoy, Hausman, Reiss and, in addition, Pearl. I focus here on Pearl since that is what he has done. It queries what counterfactuals *say* and what they are good for. Happily, for Pearl the answer to the first is clear. Suppose a situation can be represented with a set of Pearl-type equations, with values or probabilities specified for the external variables. The antecedent of the counterfactual plus the do-semantics dictates what changes occur in the equations.¹³ The new equations dictate new probabilities and values, including the counterfactual consequent. So it's a great predictive tool – so long as the first equations are correct and the changes in the situation are those represented in the antecedent and do-semantics. Unlike Pearl, I think we make far more kinds of change and in far different kinds of systems than he can handle, including ones that change the underlying structure, ala Lucas. This is my second problem with focussing on modularity: it's of very limited use.

My point is obviously not that we should be able to predict 'when there is not enough information' to do so, but rather that Pearl's methods are limited in the counterfactuals they can evaluate. And often there are other methods that work where

¹³ Note that I don't say, here or in the passage quoted from me, that the antecedent must be atomic. I apologize if what I wrote can be read that way.

Pearl's won't. After all, Lucas did not just *claim* the Phillips curve breaks down when inflation is manipulated; he produced a rational-expectations model to *show* it. And I have offered ways to model situations where causes act probabilistically and CMC breaks down, models from which it is equally possible to derive what results from specified changes. Also, ala my lemonade-biscuit model, a Hoover-style model¹⁴ can give correct predictions about real manipulations without laying out the (mechanical) causal laws, as can models in which manipulated variables are superexogenous even if not causal by other standards. Or – *HC&UT* talks about 'implementation-neutral' counterfactuals. We might want a causal connection to hold under 'any' implementation, without knowing what those will be, as in sending products where users live very differently. Often we have good reasons to suppose a counterfactual is, or isn't, true, given an implementation-neutral interpretation. But this can't be extracted by do-semantics.

HC&UT focuses explicitly on counterfactuals that provide direct predictions of what will happen when we act. Here Pearl embraces a different kind: mere conceptual changes in a representation scheme.¹⁵ Fine, but what are these counterfactuals good for? Not testing, which I have stressed is a nice thing that modularity under real changes can secure. A case Pearl defends is $P(\text{cancer}/\text{do}(\text{smoking}))$. For decades I have called this quantity 'the strength of smoking's capacity to affect cancer' and only x pages from Pearl's quote from me, I argue contra Heckman that it makes sense even where manipulating smoking by itself is not actually possible. So I admit the importance of this quantity. But I add a warning. This quantity may *measure* the

¹⁴ As I characterize Hoover. See note 7.

¹⁵ For instance, following Pearl's suggestions in equations 1)-4), in * substitute new values for p_t without changing θ . This is not what the government can actually do in the long run according to Lucas.

contribution of a ‘stable biological characteristic’ but evidence that there *is* a stable contribution must come from elsewhere.¹⁶ We can extract $P(x/\text{do}(y))$ for any x and y from Pearl-type equations governing a situation¹⁷ but this does not mean this quantity has significance for other situations. This is why I stress that ‘capacity’ is an especially strong metaphysical notion, beyond that of ‘causal law’, and knowing that a factor has a capacity requires more evidence, different in kind, than is required for causal-law claims.

$P(x/\text{do}(y))$ nicely illustrates the theme of hunting and using causes. We extract this quantity from Pearl-equations with do-semantics. We get those equations by econometric modelling plus, say, my sufficiency tests for a causal interpretation.¹⁸ But to use it in the way Pearl and I suggest, we still need to establish a great deal more. My concern is to see to it that our methods for doing so are appropriate to our uses.

¹⁶ As must the information about how this contribution combines with contributions from other causes. Contributions in mechanics add vectorially, but not those in economics. *HC&UT* describes a number of different schemes from physics and from economics for combining contributions.

¹⁷ So long as the expression is well-formed.

¹⁸ Plus an assumption of an original causal ordering on the variables that we haven’t discussed, and an assumption about the general functional form, also not discussed here.