Cartwright on Causality: Methods, Metaphysics, and Modularity

Daniel Steel

Department of Philosophy

503 S Kedzie Hall

Michigan State University

East Lansing, MI 48824-1032

USA

Email: steel@msu.edu

1. Introduction.

Nancy Cartwright's most recent book, *Hunting Causes and Using Them: Approaches to Philosophy and Economics* (hereafter, HCUT), is a welcome and provocative addition to the current literature on causation. In HCUT, Cartwright further develops themes from her earlier work, especially *Nature's Capacities and their Measurement* (1989) and *The Dappled World* (1999). One theme is that methodological issues having to with inferring and applying claims about cause and effect must be considered in tandem with metaphysical questions about what causation is. And with regard to the latter issue, Cartwright insists that causation is not just one kind of thing but is instead a general category for various types of processes that often differ in important ways. From these two themes, it naturally follows that one should be skeptical that there is any method of causal inference that is applicable in all cases. Moreover, for any method, one ought to be very clear about the types of causal systems for which it is suited and, of equal importance, those for which it is not. Given Cartwright's approach, such investigations will require careful attention to domain specific detail about the nature of the causal processes of interest. Cartwright pursues these ideas in the context of critical examinations of current approaches to causation, including Bayes nets and several approaches proposed by econometricians.

I am quite sympathetic to Cartwright's overall perspective on causation, but I take issue with some of her characterizations of particular approaches and several of her specific claims about their limitations. I focus on Cartwright's claims concerning methods of causal inference that rely on Bayes nets, which among the methods she discusses is the one I know best. First, I argue that Cartwright's discussion of this topic

is problematic insofar as it does not pay adequate attention to the distinct projects that might be pursued within a Bayes nets approach to causation. Some authors in this genre (e.g. Spirtes, Glymour, and Scheines 2000) are primarily concerned with *inferential projects*, that is, exploring what can be learned about causation from data and what predictions can be derived from causal knowledge once it is learned. In contrast, others (e.g. Woodward 2003) pursue *semantic projects*, that is, articulating a concept of causation consistent with a variety of methods of causal inference, including Bayes nets. Of course, Cartwright is correct to insist that inferential and semantic projects are interconnected. But as a result of paying inadequate attention to the difference, Cartwright sometimes misinterprets these approaches in important ways and presents complementary accounts within a common framework as if they were distinct methods. Thus, a person would be very likely to have an inaccurate view of Bayes nets approaches to causal inference if HCUT were her only source of information on that topic.

In addition, I disagree with a number of claims Cartwright makes about the limitations of Bayes nets as a method of causal inference. Although Cartwright focuses on genuine challenges, I argue that she tends to exaggerate the extent to which those challenges pose real problems. Among the concerns about Bayes nets raised by Cartwright, I will focus on those relating to something known as *modularity*. Put roughly, a causal system is modular when it is possible to alter causal relationships in one part of the system while leaving causal links elsewhere unchanged. Cartwright suggests that modularity is usually not a reasonable assumption to make about causal systems. For example, chapter 7 of HCUT is titled, "Modularity: it can—and generally does—fail." However, I argue that Cartwright's arguments against modularity consist of examples

that either are not genuine counter-instances or which show only that exceptions to modularity are possible in principle.

2. Inferential versus Semantic Projects.

In chapter 2 of HCUT, Cartwright examines what she calls "dominant accounts of causation," and Bayes nets methods are the first on her list (p. 12). According to Cartwright, Bayes nets methods presume that causal relationships are defined by reference to probabilities in a manner similar to Patrick Suppes' (1970) probabilistic theory of causality (pp. 11, 43). The subsequent chapter of HCUT discusses methods for acquiring and applying causal knowledge, and among these is an approach that Cartwright calls "invariance methods" (p. 33). In this section, I suggest that this characterization of Bayes nets and invariance approaches is mistaken in several respects. First, Bayes nets methods do not assume a probabilistic definition of causation. Indeed, the most prominent advocates of Bayes nets methods favor (implicitly or explicitly) a manipulationist rather than probabilistic theory. Secondly, the invariance approach to causation is most closely associated with Jim Woodward (2003) who is quite clear that his project is a semantic rather than an inferential one. Woodward's invariance account aims to explicate the intuitive notion that causal relationships are distinguished from mere correlations in virtue of indicating effective means of manipulation and control. Thus, rather than being a separate "method" from Bayes nets, Woodward's proposal is best understood as a complementary proposal about how the causal relationships in Bayes nets should be interpreted.

Probably the two most commonly cited texts on Bayes nets approach to causation are Spirtes, Glymour, and Schienes' (2000), *Causation, Prediction, and Search*,[1] and Pearl's (2000), *Causality*. The official line of Spirtes, Glymour, and Scheines is agnosticism about philosophical theories concerning the nature of causation (cf. Glymour 1997, 317–318). However, motivating discussions associated with their work clearly indicate the central emphasis they place on implications for manipulation, control, and policy decisions as characteristic aspects of causal relationships. For example, consider these passages from the preface to *Causation, Prediction, and Search*.

> This book is intended for anyone, regardless of discipline, who is interested in the use of statistical methods to help obtain scientific explanations or to predict the outcomes of actions, experiments or policies. … [T]he most urgent questions about the application of statistics … concern the conditions under which causal inferences and predictions of the effects of manipulations can and cannot reliably be made … (2000, p. xiii)

Likewise, consider this passage from the preface written by Glymour from a volume he co-edited on issues relating to Bayes nets approaches to causal discovery.

> *Problems of causal prediction* concern how to reliably predict the changes in some features of a kind of system that will result if an intervention changes other features of the system. This book is about the problem of learning to make causal predictions, which is the problem of causal discovery. (Glymour and Cooper 1999, xi; italics in original)

[1] Many citations of this work are of the first (1993) edition.

Although statements of this sort do not amount to a full blown analysis of causation, they pretty clearly express the judgment that, whatever causation is, knowledge of it often indicates means for manipulating one's environment to attain ends. The link between causation and manipulation is expressed even more explicitly in Pearl's *Causality*. Like Spirtes, Glymour, and Scheines, Pearl stresses the link between causation and manipulation (cf. 2000, p. 337). In addition, Pearl explicitly rejects the idea that causation can be defined by means of probabilities, writing that "the word *cause* is not in the vocabulary of probability theory; we cannot express in the language of probabilities the sentence, *mud does not cause rain*" (p. 342; italics in original). Moreover, the formal machinery presented in both of these texts addresses the link between causation and manipulation.[2]

Thus, the Bayes nets approach to causal inference elaborated in Spirtes, Glymour, and Scheines (2000) and Pearl (2000) is most naturally associated with a theory of causation in which manipulation is the central driving concept. Woodward (2003) develops just such a theory, according to which invariance under intervention is the distinctive feature of causal generalizations. Invariance is related to manipulation because generalizations that are invariant in Woodward's sense indicate variables that can, at least in principle, be used to manipulate other variables. Woodward's definition of intervention (2003, p. 98) plainly follows discussions found in Spirtes, Glymour, and Scheines (2000, pp. 48–49) and Pearl (2000, pp. 70–72).[3] In addition, connections

[2] See Spirtes, Glymour, and Scheines' manipulation theorem (2000, pp. 47–58) and Pearl's "do calculus" (2000, pp. 85–89).

[3] Pearl describes two distinct concepts of intervention: one in which the intervention sets the targeted variable to a definite particular value and the other in which the intervention itself is a variable that imposes a new probability distribution on the targeted variable (see Pearl 2000, sections 3.2.1 and 3.2.2,

between probability and intervention discussed by Woodward (cf. 2003, section 2.3) are simple consequences of his definition of intervention within the standard Bayes nets framework. Finally, Woodward is quite clear that his project is a semantic one: he aims to suggest an improved and clarified interpretation of causal claims as they occur in a variety of contexts in contemporary science. In sum, Woodward's invariance account of causal explanation provides an in depth explication of a perspective on causation apparently favored by some prominent advocates of Bayes nets approaches to causal inference.

What reason could there be, then, for claiming that Bayes nets assume that causality is defined in terms of probabilities? Cartwright cites an essay by Wolgang Spohn (2001) that develops a probabilistic definition of causality in connection with Bayes nets. However, this shows at most that a probabilistic definition of causality is consistent with a Bayes nets approach, not that it is the only option.[4] After all, the official agnostic line of Spirtes, Glymour, and Scheines suggests that several conceptions of causation might be compatible with their approach. Another possible reason has to do with propositions about the probabilistic implications of causal claims commonly assumed in by Bayes nets approaches to causal inference. In particular, the causal Markov and faithfulness conditions specify relations between causation and probability that are also presumed by probabilistic theories of causality, such as Suppes' (1970). However, propositions about the connection between probability and causality can be treated as methodological principles rather than axioms that are true solely in virtue of the

---

respectively). Woodward and Sprites, Glymour, and Scheines use the second of these two intervention concepts.

[4] For another heterodox proposal about the interpretation of causation best associated with Bayes nets see Williamson (2005).

meaning of the word "cause." And that is in fact how the causal Markov and faithfulness conditions are typically regarded. Of course, one might have doubts about the scope of application of these methodological principles, but that is a separate question from whether Bayes nets approaches to causal inference are inherently tied to a probabilistic definition of causality.

3. Modularity

Modularity is closely associated with manipulation. Roughly put, modularity states that it is possible to intervene at one place in a system without altering causal relationships elsewhere. The opposite of a modular system would be a completely holistic one in which any intervention in one place reconfigures the causal mechanisms everywhere else. Modularity is a useful feature for a machine to have, since it allows one to repair a malfunctioning component without compromising the functioning of other components. For example, one can easily imagine how difficult auto-repair would be if, say, it were impossible to replace a radiator without altering all of the other mechanisms of the engine. This intuitive idea also suggests an evolutionary argument for why modular mechanisms would be favored by natural selection: such mechanisms facilitate quick adaptations to changing environments (cf. Steel 2007, chapter 3).

A good deal of Cartwright's argument against modularity is directed at attempts to show that modularity is an essential aspect of the concept of causation. For example, she critiques Hausman and Woodward's (1999, 2004) arguments that modularity is inherent in the manipulationist conception of causation and that the causal Markov condition is a consequence of modularity. In contrast, Cartwright regards modularity as

an “epistemically convenient” (p. 81) circumstance that may facilitate manipulation when present but which has no claim to being a sine qua non of causality. I agree with Cartwright about the futility of efforts to argue for some necessary conceptual link between modularity and causality. Not only is it very difficult to establish any such connection, an absolute conceptual link between modularity and causality would not do the work we need even if it existed. Modularity is of interest because modular systems are typically less difficult to manipulate than non-modular systems. Since we frequently want to manipulate things our environment, it matters whether the systems we are interested in—economies, machines, living organisms, ecosystems, etc—are modular. Yet an absolute conceptual connection between modularity and causation would not show that this is the case: it would only show that, if a system is not modular, then it cannot be properly labeled “causal.” To put the matter another way, we cannot make the world easier to manipulate and control by insisting that nothing is causal unless it is modular. So I agree with Cartwright that, instead of dubious claims about the essence of causality, discussions of modularity should focus on the following sorts of questions. How can the intuitive notion of modularity be defined more precisely? What inferences can modularity license under what circumstances? When is modularity a reasonable assumption and when is it not?

My disagreement with Cartwright chiefly concerns her answer to the last of these three questions. On the basis of several examples, Cartwright suggests that modularity is generally not a reasonable assumption to make—that it obtains only in narrow range of unusual circumstances. I found three concrete examples in HCUT used to make this case: the carburetor (p. 15), the toaster (p. 85), and a lemonade and biscuit making

machine at the Institute for Advanced Study in Bologna (p. 209). If modularity fails for such simple everyday machines, then there would indeed be good grounds for skepticism about modularity in general. However, I do not think that any of these cases are genuine counterexamples to modularity. In the case of the carburetor, Cartwright points out that a number of the causal relationships depend on the geometry of the chamber into which air and fuel are fed. Hence, an intervention that altered the geometry of the chamber would be non-modular. But this only shows that there are *some* interventions on one feature of the system that would alter causal relationships elsewhere in the system. It is almost always possible to intervene in an indiscriminate manner. For example, if smash my computer with a hammer, I have performed an intervention that has simultaneously modified nearly all of the causal relationships that had hitherto obtained in it. The point is that modularity does not require that *all* possible interventions affect the system in a precise and focused way; it only requires that it be possible to perform *some* intervention of this sort. Cartwright has not shown that it is impossible intervene on the carburetor in a manner that alters one causal relationship without changing the others. She has only pointed out that any such intervention cannot work by changing the geometry of the chamber. Yet a carburetor has a number of other parts that could be subject to intervention: an air filter, a choke valve, a throttle valve, etc.

Cartwright's toaster example fails to be a counterexample to modularity for a somewhat different reason. The causal relationships in the toaster are represented in the graph in figure 1. Cartwright claims that the toaster is not modular because the motion of

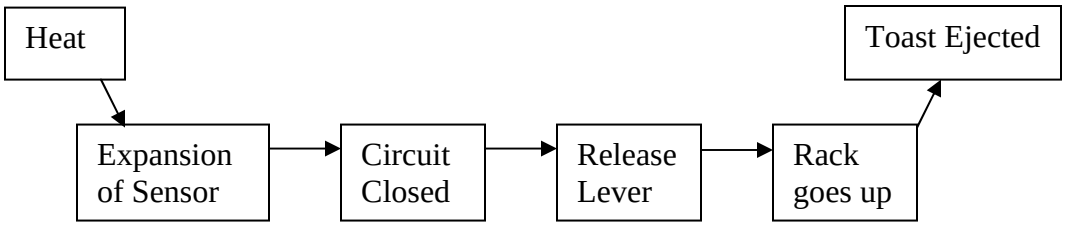

Figure 1

the rack is completely determined by the lever (p. 85). Since the rack is bolted to the lever, if the lever moves, so too must the rack. However, I think it is pretty clear that this is not a genuine counterexample to modularity and that there is little difficulty in seeing how an intervention could target the rack without changing the relationships in the causal chain from Heat to Release Lever, or the link between Rack goes up and Toast Ejected.

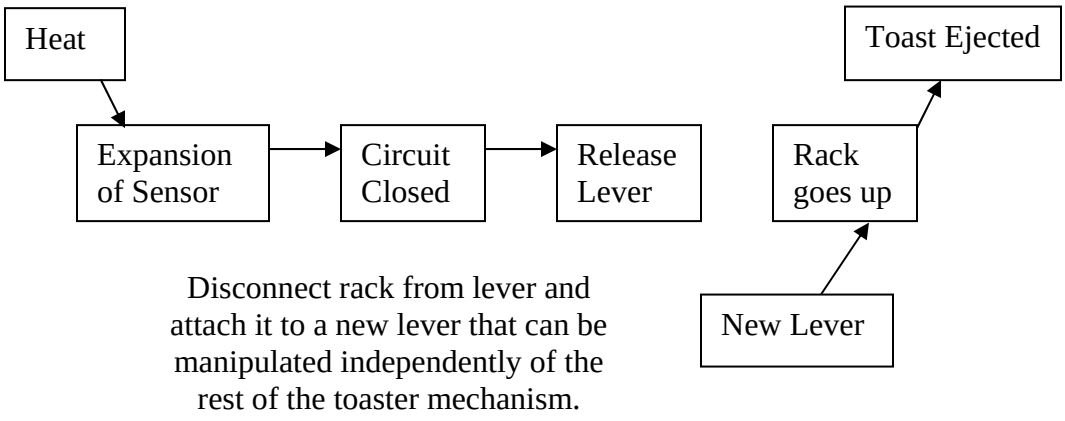

Figure 2

The graph in figure 2 explains how this could work. It would, I think, be fairly trivial for a mechanically inclined person to carry out this intervention. Cartwright's example of the lemonade-biscuit machine (pp. 209–210) is similar. This machine contained both a pump and a motor, and when the pump began to operate it would trigger the motor to

start, too. There were two levers to start the pump: one which only started the pump and second which started the pump while simultaneously damping (though apparently not stopping) the motor. The thought is apparently that modularity fails in this example because there is no lever that directly affects the motor alone. But again it would presumably require no great feat of engineering to install a third lever on the lemonade-biscuit machine that started the motor but which had no connection to the pump.

These two examples illustrate that interventions often involve introducing a new cause of a targeted component while decoupling that component from its causes within the system. Thus, a system is modular when it is possible to perform localized interventions on it, that is, to introduce new causes of a component and disconnect it from other causes without changing mechanisms elsewhere in the system. In contrast, in her discussion of the toaster and lemonade-biscuit examples Cartwright presumes that an intervention must work solely with the pre-existing mechanisms. Consequently, her thought seems to be that modularity requires that causal systems come, in effect, ready made with levers that allow one to separately manipulate each component. But that is just not what modularity asserts. Modularity is not a claim about how causal systems *are*; it is a claim about how they *can be modified*.

Of course, one can *conceive* of the abstract possibility of a machine in which it is impossible to change the existing mechanism in any way or impossible to introduce any new causes of the components. A machine of this sort, if one existed, might indeed be non-modular. At some points in HCUT (cf. pp. 211-212), Cartwright's objection to modularity seems to be only that this possibility can be imagined. But there is little reason to be concerned with a mere possibility of this sort so long as *it generally is*

possible to alter mechanisms and introduce new causes in systems we are concerned with. And indeed, it is clear that it is in fact possible to do this for a wide range of cases, including machines, organisms, and social systems. In general, when causal relationships depend on a contingent arrangement of components, it is nearly always possible to alter mechanisms (e.g. by rearranging components) or to introduce new causes (e.g. by inserting new components that are linked to the old ones). Nevertheless, I should emphasize that I am not asserting that modularity is a completely unproblematic and universally satisfied condition. For example, I think that it is an interesting question what basis there is for presuming modularity in biology and I also think that several classic arguments against the existence of laws of social phenomena attempt to show, in effect, that social mechanisms are not modular.[5] What I have argued above is only that Cartwright has provided no good reason to be skeptical of modularity.

I had a similar reaction to several other claims that Cartwright makes about the limitations of Bayes nets approaches to causal inference. That is, I agreed that she raised a genuine issue, but I felt that her arguments did not justify her pessimism. For example, Cartwright (pp. 68–72) discusses a theorem in Spirtes, Glymour, and Scheines that parameterizations that violate the faithfulness condition are contained within a subset of Lebesgue measure of the entire space of possible parameterizations.[6] I think that Cartwright is certainly correct that it needs to be explained how this theorem is supposed to provide a reason to accept the faithfulness condition. But several prior publications have attempted to fill that gap, sometimes while directly responding to the arguments against the faithfulness condition presented in HCUT (cf. Pearl 1998; Woodward 1998;

[5] For further discussion of these issues, see chapters 3 and 8 of Steel (2007).
[6] See Spirtes, Glymour, and Scheines' theorem 3.2 (2000, pp. 41–42).

Steel 2006). To provide just one further example, Cartwright complains that some Bayes nets methods presume causal sufficiency, that is, that there are no unmeasured common causes (pp. 74–75). But she does not mention that discovery algorithms also exist in the Bayes nets literature that do not rely on this assumption (cf. Spirtes, Glymour, and Scheines 2000, chapter 6; Pearl 2000, section 2.6; Neopolitan 2004, section 8.5).

4. Conclusion

My general assessment of HCUT, then, is sympathy towards its central themes but skepticism about a number of its particular claims. But even where I disagree, I find that Cartwright plays the valuable role of a Socratic gadfly. She draws attention to underlying principles of popular approaches to causal inference and insists that clear arguments be provided for them and that these arguments be tied to the circumstances in which the methods are intended to be applied. Her work, therefore, is a powerful antidote to the complacency and aversion to thinking outside the box that are common undesirable side effects of a successful research program. Those interested in current theories of causation owe a debt to Cartwright's indefatigable questioning and refusal to merely accept what is commonly assumed.

References

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

_____ (1999), *The Dappled World: A Study of the Boundaries of Science*. Cambridge: Cambridge University Press.

_____ (2007), *Hunting Causes and Using Them: Approaches in Philosophy and Economics*. Cambridge: Cambridge University Press.

Glymour, C. (1997), "A Review of Recent Work on the Foundations of Causal Inference", in V. McKim and S. Turner (eds.), pp. 201-248.

Glymour, Clark and Gregory Cooper (eds.) (1999), *Computation, Causation, and Discovery*. Cambridge, MA: MIT Press.

Hausman, Daniel and James Woodward (1999), "Independence, Invariance and the Causal Markov Condition", *British Journal for the Philosophy of Science* 50: 521-83.

_____ (2004), "Modularity and the Causal Markov Condition: A Restatement", British Journal for the Philosophy of Science 55: 147-161.

McKim, V., and S. Turner (eds.) (1997), *Causality in Crisis? Statistical Methods and the Search for Causal Knowledge in the Social Sciences*. South Bend, IA: University of Notre Dame Press.

Neopolitan, R. (2004), *Learning Bayesian Networks*. Upper Saddle River, NJ: Prentice Hall.

Pearl, J. (1998), "TETRAD and SEM", *Multivariate Behavioral Research* 33: 119-128.

_____ (2000), *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.

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

_____ (2000), *Causation, Prediction, and Search*, 2nd edition. Cambridge, MA: MIT Press.

Spohn, W. (2001), “Bayesian Nets are all there is to Causal Dependence”, in D. Costantini, M. Galavottia and P. Suppes (eds.), *Stochastic Causality*, Stanford, CA CSLI Publications.

Steel, D. (2006), “Homogeneity, Selection, and the Faithfulness Condition”, *Minds and Machines* 16: 303-317.

_____ (2007), *Across the Boundaries: Extrapolation in Biology and Social Science*. New York: Oxford University Press.

Suppes, P. (1970), *A Probabilistic Theory of Causality*. Amsterdam: North-Holland.

Woodward, J. (1998), “Causal Independence and Faithfulness”, *Multivariate Behavioral Research* 33: 129-148.

_____ (2003), *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

Williamson, J. (2005), *Bayesian Nets and Causality: Philosophical and Computational Foundations*. Oxford: Oxford University Press.