The Philosophical Review, Vol. 110, No. 4 (October 2001)

CAUSALITY: MODELS, REASONING AND INFERENCE. By JUDEA PEARL. Cambridge: Cambridge University Press, 2000. Pp. xvi, 384.

Judea Pearl has been at the forefront of research in the burgeoning field of causal modeling, and *Causality* is the culmination of his work over the last dozen or so years. For philosophers of science with a serious interest in causal modeling, *Causality* is simply mandatory reading. Chapter 2, in particular, addresses many of the issues familiar from works such as *Causation, Prediction and Search* by Peter Spirtes, Clark Glymour, and Richard Scheines (New York: Springer-Verlag, 1993). But philosophers with a more general interest in causation will also profit from reading Pearl's book, especially the material in chapters 7, 9, and 10 (not to mention the delightful epilogue), which is self-contained and less technical than other parts of the book. The present review is aimed primarily at readers of the second type.

Pearl represents a system of causal relationships by a *causal model*. A causal model consists of a set of *variables*, a set of *functions*, and a *probability* measure representing our ignorance of the actual values of the variables. Each function generates an equation of the form $V_i = f_i(V_{i1},\ldots,V_{im})$, where V_i is distinct from each V_{ij}. These equations represent "mechanisms" whereby the value of one variable is causally determined by the values of others. Mechanisms differ from what philosophers call "laws" in that the former are asymmetric. If it is a law that $Y = f(X)$ (and f is an invertible function), then it is also a law that $X = f^{-1}(Y)$. By contrast, if a causal model contains the mechanism $Y = f(X)$, then it will not also contain the mechanism $X = f^{-1}(Y)$ (except in very special cases). The system of equations may be represented qualitatively in a directed graph, with an "arrow" drawn from V_i to V_j just in case V_i figures in the function for V_j. The directed graph representation greatly facilitates inferences about the model.

A causal model may be used to evaluate counterfactuals of the following form: if the value of V_i were v_i, then the value of V_j would be .... The resultant value of V_j is determined by replacing the equation $V_i = f_i(V_{i1},\ldots,V_{im})$ with $V_i = v_i$, and then solving the resulting system of equations. This replacement indicates that V_i is set directly to v_i by an intervention from outside the system, rather than having its value causally determined by the values of the variables within the system. The intervention need not be miraculous: mechanisms are not inviolable laws, but rather ceteris paribus laws that can be disrupted by external interventions. Such an intervention will *not* affect the functional forms of the other mechanisms in the causal system: the mechanisms are *autonomous*.

The bulk of Pearl's book deals with inference problems where we have only partial information about the causal system being modeled. Our partial information may be of various kinds. Observational evidence may give us information about probabilistic correlations between variables; background assumptions may give us information about the graphical structure; and con-

trolled experiments may give us information about the results of interventions. Using such information, we may want to infer unknown features of the graphical structure, or estimate the effect of a new intervention, or compute the probability of a counterfactual claim.

My concern here will not be with the details of these inferences, but rather with the picture of causation that seems to underlie Pearl's success in arriving at those details. In many ways, Pearl's approach is at odds with standard philosophical thinking about causation, and philosophers would do well to take heed.

Causal pluralism. Most philosophers talk as though there is one specific relation—causation—that is the target of philosophical inquiry. In *Causality*, one finds definitions of causal effect, causal relevance, total effect, direct effect, actual cause, contributing cause, and so on: the 'Causality' of Pearl's title does not name some specific relation, but rather an entire subject matter. Pearl's concept of actual causation comes closest to the notion of "token causation" that most philosophers take to be central. It is telling that this concept appears only in Pearl's final chapter; the concept is not needed for Pearl's treatment of rational deliberation, counterfactuals, experimental methodology, and so on.

Reduction. Pearl's conceptual toolkit includes the following two primitives: the notion of a mechanism and the notion of setting the value of a variable through an intervention. These are thoroughly causal notions: Pearl offers no reduction of the causal to the noncausal. Many philosophers assume that nonreductive accounts of causation must be trivial and viciously circular. But if there are many interesting causal relations, there is nothing trivial or circular in defining some of these relations in terms of others. For example, Pearl's definition of actual causation in terms of mechanisms and interventions is far from trivial.

Epistemology. The perceived need for a reductive analysis of causation stems, in part, from an armchair epistemological argument: we do not directly observe causal relations, so if we are to have any knowledge of what causes what, we must somehow be able to derive causal knowledge from our knowledge of noncausal facts. Pearl provides us with a much more realistic epistemology: we make causal inferences from noncausal facts *together with* defeasible background assumptions that are themselves causal.[1]

Actual Causation. Most philosophers will find the most exciting part of Pearl's book to be his treatment of actual causation. Pearl is the first writer in the causal modeling tradition to tackle this thorny issue, and the infusion of new ideas from outside of philosophy is most welcome. Pearl's account promises to accommodate causal preemption without invoking the problematic the-

[1]Pearl is not the first to try to let this cat out of the bag: see for example N. Cartwright, "Causal Laws and Effective Strategies," reprinted in her *How the Laws of Physics Lie* (Oxford: Oxford University Press, Clarendon Press, 1983).

sis of causal transitivity. Suppose that Assassin shoots at Victim, who dies. Had Assassin not shot, Backup would have shot Victim instead. An appropriate analysis of this case should yield the intuitively correct result that Assassin's shot is an actual cause of Victim's death, even though Victim's death does not depend counterfactually upon Assassin's shot. Pearl obtains this result by constructing a new model, with modified mechanisms, in which Victim's death does depend counterfactually upon Assassin's shot. Informally, Assassin's shot is an actual cause of Victim's death because it protects that outcome against the breakdown of the other mechanisms represented in the model. That is, Assassin's shot would ensure Victim's death even if Backup were distracted, his gun were jammed, and so on.

The Methodology of Modeling. I conclude by pointing to one area where the potential for engagement with philosophical issues remains undeveloped in Pearl's book. There are important methodological issues about how best to model a given causal system. These issues ought to be of interest to philosophers, in part because philosophers of science are professionally concerned with issues of scientific methodology, and in part because these methodological issues serve as useful surrogates for murkier metaphysical issues. For example, metaphysicians have worried about whether the *time* at which an event occurs is an *essential* property of that event: if a fire had occurred on Tuesday instead of Monday, would it have been the same event? It may be more fruitful to replace this question with a new one: Should we construct a model using one variable F, taking the value 1 if a fire occurs on Monday and 2 if it occurs on Tuesday; or should we use two variables F_m and F_t, taking the value 1 if a fire occurs on the subscripted day, and 0 otherwise? Just what hinges on this choice? What errors might we commit if we make the wrong choice? It would be nice to have Pearl weigh in on such methodological questions.[2]

CHRISTOPHER HITCHCOCK

California Institute of Technology

[2]I would like to thank Alan Hájek, Paul Humphreys, Judea Pearl, Jonathan Schaffer, and Jim Woodward for helpful comments and suggestions.