

## BOOK REVIEW

# A Structural Approach to the Understanding of Causes, Effects, and Judgment

Judea Pearl. **Causality: Models, Reasoning, and Inference.**

Cambridge, UK: Cambridge University Press, 2000. xvi+ 384 pp. \$39.95.  
ISBN 0 521 77362 8.

*Reviewed by* Susan F. Butler

---

Judea Pearl is Professor Emeritus of Computer Science and Statistics and Director of the Cognitive Systems Laboratory at University of California, Los Angeles. He has authored the books *Heuristics* (1984) and *Probabilistic Reasoning in Intelligent Systems* (1988). He has also published close to 300 articles on automated reasoning, learning, and inference. In 1999, he was awarded the IJCAI Research Excellence Award in Artificial Intelligence fundamental work in heuristic search, reasoning under uncertainty, and causality.

Susan F. Butler is a doctoral student at Tufts University. Her work is in modeling of judgment and decision making, stochastic processes, and the examination of choice behavior under uncertainty. © 2002 Elsevier Science (USA)

---

In his recent book, *Causality*, Judea Pearl develops further ideas and applications he discusses in his prior books, *Probabilistic Reasoning in Intelligent Systems* (1988), and *Heuristics* (1984). His goal of developing a positive alternative to counterfactual null hypothesis testing is mostly met. However, he has fallen short of his goal of developing a logical and mathematical procedure by which judgment can be automated as a heuristic applicable to any context. The risk is that causality determined by machine and heuristics may yield very unlikely conclusions.

In Pearl's opinion, traditional methods for investigating probabilistic relationships evaluate only surface phenomena and fail to describe the structure of the causal relationships determining the observed behavior. For example, instead of specifying counterfactual formulations as with null hypothesis statistical tests, Pearl thinks it better to specify individual variables and the nature of their relationships. The directionality and probability between each pair of variables can be specified, creating a graphical map and a means of testing, specifying, and developing the

Address correspondence and reprint requests to Susan F. Butler, Psychology Department, 490 Boston Avenue, Medford, MA 02155.



structure of the graphical causal model. His distinction between observed surface phenomena and underlying causal relationships is a useful and important one.

The review in Chapter 1 of both classical and Bayesian probability is simple and clear. Readers unfamiliar with Bayesian methods are provided with a basic framework. For those readers familiar with this material, it is easily passed over. Central to Pearl's argument is the graphical notation of trees, chains, and acyclic structures used to represent a complete mapping of the independent, intervening, confounding, and dependent variables that are required to describe the causal structure of an event. He does an excellent job of communicating these basic tools.

Pearl postulates that dependency is caused by the flow of information. The graphical maps of the dependencies between variables describe both the specific connections and the directionality of the information flow among the variables. If there is no flow of information, there is no effect of one variable on another and no dependency. He differentiates between related variables and unrelated variables by using the process of *d-separation*. D-separation is established when one variable blocks the flow of information, or causal connections, between two other variables.

As an example of dependency, flow of information, and d-separation, Pearl explains Berkson's paradox, or selection bias. He writes "... observations on a common consequence of two independent causes tend to render those causes dependent because information about one of the causes tends to make the other more or less likely, given that consequence has already occurred" (Pearl, p. 17). Here, the flow of information is used to formulate the structure of dependency. D-separation would block the flow of information in the graphical model or interrupt the dependency between variables.

Pearl contends that the proposed method is an improvement over null hypothesis testing because simple statistical association does not imply causality. Pearl's method provides a systematic form for consideration of two or more variables in a variety of relationships. If the predictions do not fit the data, the predicting variables and their relationships must be reconsidered. Perhaps existing relationships between variables are inaccurate. Perhaps some confounding variable has been omitted. The model is not one of a singular cause and a singular effect, but rather one of causes and effects.

Data are used to test the validity of the proposed structure. If the predictions of the model do not fit the data, the given map does not explain the observations. The variables and their relationships can be reevaluated to reveal previously unrecognized intervening variables or misstated relationships. Both the variables and their relationships can be reformulated until a model fits. The structure of causality is thus dissected into multiple specified variables and their relationships, or, as Pearl says, he is "mathematizing causality."

While the goal is grand, and the rigor of the proposed logical structure thorough and well reasoned, his thesis is unlikely to have implications as far reaching as Pearl hopes. The graphical models are clearly related to systems analyses, but better developed, and greatly enriched by their association with Bayesian belief networks.

Pearl acknowledges the importance of not overdeveloping the model. He cites Ockham (1349) and Cartwright (1983, 1989, 1995a, 1995b, 1999). His goal is to keep logical embellishments to a minimum by making two assumptions: minimality

and stability. The model must be minimal, using the least complex model structure to describe the observed data. The model must also be stable, yielding output that is invariant over varying conditions. Both the descriptive and the predictive capacities of the model must be constant across varying conditions.

The effect of time on the perception and conclusion of causality is a classic matter for discussion. Simple temporal order is a criterion often used: causes precede effects. Contemporaneousness and correlation are often common confounding attributes in the understanding of cause and effect. Pearl believes a causal model that can “generalize from one behavior under one set of circumstances to behavior under another set of circumstances” (p. 61) offers a better test of causality. Temporal order, contemporaneousness, and covariance offer less, as they often have confounded variables and spurious associations.

Pearl also formulates the notion of statistical time, which is different from physical time. Generally, we believe causes precede effects. However, when considering conditionality across time, we find inconsistencies in our application of conditionality. When we evaluate the past and present, we believe there is a conditional dependency. The present is caused by factors in the past. Yet when we evaluate the present and the future, they are often considered conditionally independent. We are less willing to say what the present will bring about in the future. Pearl’s example of multivariate, economic time series is an interesting discussion of this problem. He concludes that statistical time, invariant over varying conditions and multisituationally stable, offers a better basis for conclusions of causality, particularly for machine learning applications in economics and data mining, i.e., programs designed to find stable causal relationships among variables in a given data set. This use may be more appropriate to applied contexts than to scientific inquiry.

Having considered the effect of different time contexts on the modeling of causality, Pearl also considers the effect of point of view on the modeling of causality. How do expectation, utility, and point of view affect the functioning of the model? He makes a distinction between an action, the internal decision-making process, and an act, the behavior observed from the exterior. He defines the  $do(\cdot)$  function as a function defining the act. This difference between action and act is one of locus of control or point of view. Pearl believes statisticians focus their analyses on acts or the observed behavior. In contrast, he sees economists as interested in actions and the internal decision-making process.

His discussion of the analysis of dependencies and control of confounders to maximize accuracy in the structure of the dependencies is well considered. His concept of identifiability defines the conditionality and the dependence generated by the flow of information, making latent variables and confounds clear. By conditionalizing on the  $do(\cdot)$  function, and using the graphical map, the proposed causal structure can be tested. Using the criterion of identifiability gives a means of determining if one variable is truly dependent on another variable, even in the context of many variables.

Expanding the contexts in which causal relationships can be discerned, Pearl describes modeling and analysis using both experimental and nonexperimental information. Causal diagrams give a mathematical tool to explain findings as well as a structure with which to query assumptions as to their adequacy to identify

causal relationships and effects. By using a formal mathematical language and structure, one can test assumptions underlying the model and its predictive capacities. The examples are well presented and discussed; they refer nicely to some classical works in the field over the past century. The range of reference throughout this book is excellent.

Pearl evaluates the uses of structural equation modeling in both economics and the social sciences. In Chapter 5, he reviews the history of these methodologies and describes the problems these approaches have had in drawing conclusions about causality. He discusses error terms, decomposition of effects, and the differentiation of exogenous from endogenous variables in economic models. He believes economists use them as density functions, whereas, in contrast, social scientists use them as summaries of covariance matrices. Neither group uses structural equation modeling as a means of describing causality. Pearl claims structural equation modeling does not yield causal conclusions. He proposes that it could be used in combination with graphical modeling to map and measure causality accurately and unambiguously in any domain.

Aiming to improve the situation, his goal is to see the wide use of a systematic method of managing models with unmeasured confounders to generate a structure that moves away from counterfactual independencies toward processes and mechanisms establishing causality and more closely approximating human judgment. The application of causal diagrams to dynamic process control resembles systems analyses. The systems structure is used to identify causal quantities relative to the model as distinct from statistical parameters. There is a risk that several models may be supported by the same data. This ambiguity can be resolved by using the graphical model and computation of each interaction and variable. It is the quantification of the variables and their interaction in the context of the overall graphical model that gives us a new means of specification and testing.

In Chapter 6, Pearl states his goal simply as "... to see problems involving the control of confounding reduced to simple mathematical routines" (p. 173). This kind of simplification is a goal of major proportions. He tackles problems of confounding, collapsibility, and Simpson's paradox. His discussion of this paradox is thorough and detailed, starting with its first discussion by Pearson in 1899. He resolves it by suggesting his "sure-thing principle," that any action increasing the likelihood of an event in the subpopulations must also increase the likelihood of it in the overall population, assuming it does not change the distribution of the subpopulations. He observes that we often confuse rates and proportions. Although the discussion of this paradox is interesting, it does not reduce the control of confounding to a simple mathematical routine.

Throughout *Causality*, many examples of both familiar and unfamiliar causal relationships are used. The examples are helpful, but also problematic. They are demonstrations of problems already solved. Pearl's dream is of a program of heuristics that could simply provide you with the answer to the questions, "What is the cause? What are the many causes? How are they interrelated?" Nonetheless, he does not have a device into which you can put data, out of which will come well-sorted causes and effects. To succeed, he would have to be able to take a current unsolved problem from physics, economics, or cognition and correctly produce the

structure of independent, dependent, confounding, and intervening variables. This would be quite a test to pass. He doesn't, but he does advocate the ideal of a direct pursuit of well-specified causality, and a method of testing all the pertinent variables. Always, it is up to the researcher to determine what is to be studied and how the variables are to be measured and understood in the structure of dependency.

Confounding and no-confounding are considered specifically with regard to sufficient criteria for no-confounding and necessary criteria for confounding. Graphical causality models describing this distinction are presented and epidemiological applications are discussed. Confounding and collapsibility are re-identified as distinct, and neither is implied by the other. Pearl's goal is to make these methods of modeling and analysis easily usable by the ordinary researcher. This lofty goal might be accomplished if the format of this book resembled his book *Probabilistic Reasoning in Intelligent Systems*, where he had specific examples and exercises in the text. Were he to develop a computer program that could take input, pose questions, and develop graphical models of causality, it would likely facilitate this learning and utilization by the rank and file researcher. On his Web site, Pearl has added helpful visuals and homework that can be used in the teaching of the material.

In Chapter 7, surprisingly toward the end of the book, Pearl discusses the logical structure of counterfactuals in their various configurations and their relationship to the graphical mapping of causality. This discussion might have been better included in his early arguments, as it is out of the context of counterfactual logic that his arguments have sprung. The terms and methods of counterfactual logic would have been familiar and useful in orienting to his approach. The examples he uses of a firing squad, an econometric analysis of demand-price equilibrium, and of Ohm's law are varied, familiar, and excellent.

The functional mapping of causal order is derived from a series of successive tests for causal directionality. By testing whether variables are background or endogenous, and by testing the overall configuration of the model, a functional model for directionality is determined. Axiomatic descriptions offer a means of testing equivalence or subsumption of different systems as alternative formulations. Composition, effectiveness, and reversibility are three axiomatic properties of counterfactuals which when considered with axiomatic tests of causal relevance are useful tests of causality. The economic methods are well thought through, particularly for tests of exogeneity and endogeneity. The demonstration of exogenous variables is useful in both economic and policy analyses.

Oddly, at this point, Pearl begins a discussion about how people learn about causal relationships and how it is learned developmentally. He uses qualitative examples of children learning about causality by manipulation and testing as well as explanation. The lack of a computational approach for this material is in sharp contrast to the rest of the book.

The discussion of the risk of context shifts and reversals of the direction of causality or circularity is incomplete. In the discussion of structural versus probabilistic causality, the decision as to the accuracy or relevance of contextual factors is described as central to the success of the causal model. Yet at the end of that chapter (Chap. 7) Pearl writes "The original dream of rendering causal claims

testable was given up in the probabilistic framework as soon as unmeasured entities (e.g., state of the world, background context, causal relevance, susceptibility) were allowed to infiltrate the analysis, and methodologies for answering questions of testability have moved over to the structural-counterfactual framework. ... The ideal of remaining compatible with the teachings of nondeterministic physics seems to be the only viable aspect remaining in the program of probabilistic causation” (p. 257). He questions this goal, suggesting that “probabilistic causality” should be reassessed, and recommending that “artificial intelligence and cognitive science ... use robots programmed to emulate the quasi-deterministic macroscopic approximations of LaPlace and Einstein [likely to] outperform those built on the correct but counterintuitive theories of Born, Heisenberg, and Bohr” (p. 257). This approach seems relativistic and rash in comparison with Burnham and Anderson’s (1998) rigorous discussion of model selection, information, and appropriate level of model fitting.

The problem of the analysis of data that is less than perfect in its structure is comprehensively treated in an entire chapter. After a thoughtful discussion of the various ways data can be flawed, and the ethical causes and implications of those variations, he proposes a means of managing the situation. By determining the *average causal effect*, one can estimate, or bound, the expression for the observed probabilities. The upper and lower bounds can be optimized for the “tightest assumption-free bounds on the quantities involved” in a method resembling the Bayesian highest density interval.

In order to consider the effect of the removal of a treatment on a population, the same method can be applied to the analysis of the impact of the treatment on the treated rather than on the untreated. “Under the conditions of *no intrusion*” (Pearl, 2000, p. 269), the natural bounds define a case where a variable like a treatment or policy is removed. This is very useful in policy analysis as well as in clinical trials. The methodology is applied to a number of real world situations and methodological problems.

With regard to the structure of legal proofs, the requirements of necessity and of sufficiency are considered, from a traditional counterfactual approach and from the approach of graphical mapping of causality. The requirements of identifiability, exogeneity, and monotonicity are all detailed as to the proof of causality. Applications in medicine, random processes, forensics, and economics are also included and enrich the discussion.

The final chapter is entitled, “The Actual Cause,” as if there were a simple final analysis to a book of such breadth and complexity. He challenges the approach of testing constructed chains of counterfactuals. These models do not include an accounting for latent variables. However, Pearl does not recognize that the scientist using probabilistic approaches considers latent variables in the thinking before he or she specifies the experimental variables, not in the process of data collection and testing. Pearl claims that where causes are overdetermined or mutually contributory, but not individually separable, chains of counterfactuals cannot accurately represent the causal structure. Pearl proposes using the action variable,  $do(\cdot)$ , and a new structure he introduces, “causal beams,” from which confounding variables can be pruned, meeting the requirement of minimality. Pearl believes this approach

provides “a sufficient and non-trivial explanation for each event.” The introduction of causal beams raises the question of whether his argument requires some structural support. Nonetheless, it is a useful metaphor for the decision making that the scientist must do in order to construct a valid understanding of the matter at hand. It usefully offers a model for understanding with a structure more complex than simple linearity.

As to how to decide which “facet of causation” to use in a model, Pearl recommends relying on “pragmatics.” In the final analysis, he does not have a computational basis for the mathematizing of judgment. He does not have a heuristic of judgment able to explain, revise itself, and make raw judgments of causes and effects from the data. The models all need human participation. The goal of an artificial intelligence heuristic that can make judgments from raw data is a task he leaves for “future investigation” (p. 331).

Although this book has met high standards for its logical, mathematical, and conceptual content, it does not attain the larger goal it set out to meet, to provide an artificial intelligence heuristic of judgment. There are also problems with continuity, order, and level of discourse. At times, the various arguments about one another without continuity. The order in which the material is covered could be improved. The variability of levels of discourse makes it difficult to recommend to any one group, as there will always be some content inappropriate to the capabilities or interests of the reader. The Epilogue and the nontechnical sections are fine for the beginner. The mathematics are appropriate for one with expert knowledge. It is impossible to write a book that speaks to all levels of readership. Inevitably one reader will be lost while another is quite familiar with the content. Even considering this difficulty, the book is a worthwhile read. The many thoughtful formulations of the process of determining causality are an important addition to the current literature on proof. The methodological directions he urges researchers in many fields to pursue are worth considering.

## REFERENCES

- Burnham, K. P., & Anderson, D. R. (1998). *Model selection and inference: A practical information-theoretic approach*. New York: Springer-Verlag.
- Cartwright, N. (1983). *How the laws of physics lie*. Oxford, UK.: Clarendon.
- Cartwright, N. (1989). *Nature's capacities and their measurement*. Oxford, UK.: Clarendon.
- Cartwright, N. (1995a). False idealization: A philosophical threat to scientific method. *Philosophical Studies*, 77, 339–352.
- Cartwright, N. (1995b). Probabilities and experiments. *Journal of Econometrics*, 67, 47–59.
- Cartwright, N. (1999). Causality: Independence and determinism. In A. Gammerman (Ed.), *Causal models and intelligent data management* (pp. 51–63). Berlin: Springer-Verlag.
- Ockham, W. (1974). *Ockham's theory of terms, Part I of the summa logicae* (M. J. Loux, Trans.). Notre Dame, IN: University of Notre Dame Press (Original work published in 1349).