

Review of N. Cartwright “Hunting Causes and Using Them”*

Judea Pearl

Cognitive Systems Laboratory

Computer Science Department

University of California, Los Angeles, CA 90024 USA

September 10, 2008

(To appear in *Economics and Philosophy*.)

1 Introduction

Nancy Cartwright’s recent book “Hunting Causes and Using Them” comes to causality in an interesting stage of her stormy, century-old courtship with economics. A survey article by Kevin Hoover (2004, “Lost causes”) counts the frequency of causal terminology in econometrics articles and asks, “Where have all the causes gone?” Hoover notes a steady decline from the 1930’s to the early 1980’s, a period of over half a century, followed by two decades of sluggish recovery, to be followed by a marked upsurge of usage in the year 2000. This pattern is both enigmatic and pathological considering that the central aim of econometrics is to provide a methodology for understanding and controlling economical phenomena, and that the founding fathers of econometrics were the ones who developed the basic mathematical tools for causal analysis, in the form of structural equation models (Haavelmo (1943), Marschak (1950), and Strotz and Wold

*Portions of this review are extracted from the 2nd edition of my book *Causality* (forthcoming, 2009).

(1960)). The decline of causal understanding in econometrics has in fact been so severe, that in 1995, while I was working on my book *Causality* [Pearl, 2000a], Jim Heckman had difficulty naming a single econometric textbook that adequately treats causality.¹

Given this state of affairs, it is not surprising that philosophers of science, intrigued by the enigma, would avail themselves to offer clarity and guidance to a field that has let statistical dogma trample its glorious past. I was keenly curious therefore to read Cartwright's take on the current state of causality in economics, and to learn how she proposes to reconcile economics tradition with modern advances in causal analysis, primarily those based on causal diagrams and the logic of counterfactuals.

I was disappointed on three counts. First, Cartwright echoes, rather than leads. She tells us what economists such as Heckman, Hoover, Leroy and Hendry thought, wrote or argued, she occasionally tells us "what's wrong" with what they thought, wrote or argued, but she does not tell us "what's right," namely, how we or economists in general, *ought* to think about causation, how we should evaluate policies, how we should articulate causal assumptions, if needed, how we ought to define quantities that we wish evaluated (e.g., counterfactuals, causal effects, merits of policies) and how we ought to reason mathematically from assumptions to conclusions. Since economists admit to the chaotic state of affairs in their court, the role of philosophers should be to radiate clarity and suggest unification; by echoing diversity, we amplify hesitancy and overshadow commonality.

Second, "Hunting causes" seems to fall into some of the conceptual traps that lured economists into dead alleys and prevented the formation of a uniform consensus regarding the definition of causal concepts, their identification and their practical application. Finally, and this is naturally my main concern, Cartwright's objection to the "surgery" semantics as the basis for counterfactual and causal analysis may further distance economists from the one formalism capable of unifying their ranks and resolving their difficulties.

¹Christ (1966), long out of print, was the only one he could name. Today, thirteen years later, I doubt whether that number has increased. And if any economics professor thinks I am exaggerating, I suggest testing if his/her students can write down a formula for the sentence "treatment does not change gender," or verify whether a given treatment (or policy or program or decision) is "ignorable," or "unconfounded" or "superexogenous" in a given, fully specified economic model. For additional observations on the misunderstanding of causality in economics, see Pearl (2000a, pp. 134-138, 165-171, 215-217).

I will elaborate.

2 Cartwright objections to the $do(x)$ -calculus

Cartwright expresses several objections to the $do(x)$ operator and the “surgery” semantics on which it is based [Pearl, 2000a, p. 72, p. 201], both are essential to the definition of counterfactuals and the evaluation of causal effects.

Cartwright description of surgery goes as follows:

“Pearl gives a precise and detailed semantics for counterfactuals. But what is the semantics a semantics of? The particular semantics Pearl develops is unsuited to a host of natural language uses of counterfactuals, especially those for planning and evaluation of the kind I have been discussing. That is because of the special way in which he imagines that the counterfactual antecedent will be brought about: by a precise incision that changes exactly the counterfactual antecedent and nothing else (except what follows causally from just that difference). But when we consider implementing a policy, this is not at all the question we need to ask. For policy and evaluation we generally want to know what would happen were the policy really set in place. And whatever we know about how it might be put in place, the one thing we can usually be sure of is that it will not be by a precise incision of the kind Pearl assumes.

Consider for example Pearl’s axiom of composition, which he proves to hold in all causal models - given his characterization of a causal model and his semantics for counterfactuals. This axiom states that ‘if we force a variable (W) to a value w that it would have had without our intervention, then the intervention will have no effect on other variables in the system’ (p. 229). This axiom is reasonable if we envisage interventions that bring about the antecedent of the counterfactual in as minimal a way as possible. But it is clearly violated in a great many realistic cases. Often we have no idea whether the antecedent will in fact obtain or not, and this is true even if we allow that the governing principles are deterministic. We implement a policy to

ensure that it will obtain – and the policy may affect a host of changes in other variables in the system, some envisaged and some not” [Cartwright, 2007, pp. 246-7].

Cartwright’s objections can thus be summarized in three claims, each will be addressed separately.

1. In most studies we need to predict the effect of non-atomic interventions.
2. For policy evaluation “we generally want to know what would happen were the policy really set in place,” but, unfortunately, “the policy may affect a host of changes in other variables in the system, some envisaged and some not.”
3. Because practical policies are non-atomic, they cannot be evaluated from the atomic-semantics of the $do(x)$ -calculus even if we *could* envisage the variables that are affected by the policy.

Let us start with claim (2) – the easiest one to disprove, and by a simple argument: There is no way a model can predict the effect of an action unless one specifies correctly what variables in the model are (directly) impacted by that action, and how. In other words, under the state of ignorance described in claim (2) of Cartwright, a policy evaluation study must end with a trivial answer: There is not enough information, hence, anything can happen. It is like pressing an unfamiliar button in the dark, or trying to solve two equations with three unknowns. Moreover, the *do*-calculus can be used to verify if the state of ignorance in any given situation should justify such a trivial answer. Thus, it would be a mistake to assume that serious policy evaluation studies are conducted under such a state of ignorance; all policy analyses with which I am familiar commence by assuming knowledge of the variables affected by the policy, and expressing that knowledge formally.

Claim (1) may apply in some cases, but certainly not in most; in many studies our goal is not to predict the effect of the crude, non-atomic intervention that we are about to implement but, rather, to evaluate an ideal, atomic policy that cannot be implemented given the available tools, but represents nevertheless theoretical relationships that are pivotal for our understanding of the domain.

An example will help. Smoking cannot be stopped by any legal or educational means available to us today; cigarette advertising can. That does not stop researchers from aiming to estimate “the effect of smoking on cancer,” and doing so from experiments in which they vary the instrument – cigarette advertisement – not smoking.

The reason they would be interested in the atomic intervention $P(\text{cancer}|\text{do}(\text{smoking}))$ rather than (or in addition to) $P(\text{cancer}|\text{do}(\text{advertising}))$ is that the former represents a stable biological characteristic of the population, uncontaminated by social factors that govern susceptibility to advertisement. With the help of this stable characteristic one can assess the effects of a wide variety of practical policies, each employing a different smoking reduction instrument.

Finally, claim (3) is demonstratively disproved in almost every chapter of my book *Causality* [Pearl, 2000a]. What could be more non-atomic than a policy involving a sequence of actions, each chosen in response to a set of observations Z which, in turn, are affected by previous actions (ibid p. 75-76, 118-126). Remarkably, the effect of implementing such a complex policy can be predicted using the “surgical” semantics of the *do*-calculus in much the same way that properties of complex molecules can be predicted in atomic physics.

I have once challenged Nancy Cartwright [Pearl, 2003] and I would like to challenge her again, to cite a single example of a policy that cannot either be specified and analyzed using the *do*(x) operators, or trivially proclaimed “unpredictable” (e.g., pressing an unfamiliar button in the dark), again using the calculus of *do*(x) operators. Ironically, shunning mathematics based on ideal atomic intervention may condemn scientists to ineptness in handling realistic non-atomic interventions.

Science and mathematics are full of auxiliary abstract quantities that are not directly measured or tested, but serve to analyze those that are [Pearl, 2000b]. Pure chemical elements do not exist in nature, yet they are indispensable to the understanding of alloys and compounds. Negative numbers (let alone imaginary numbers) do not exist in isolation, yet they are instrumental in the understanding of arithmetic operations on positive numbers.

The broad set of causal problems tackled and solved in the past decade testifies that, invariably, questions about interventions and experimentation, ideal as well as non-ideal, practical as well as epistemological, can be formulated precisely and managed systematically using the atomic intervention as a primitive notion. Surely causes come in a variety of colors, including

total, direct, and indirect causes, necessary and sufficient causes, actual and generic causes but, as shown in [Pearl, 2000a], all can be analyzed and understood within a single formal framework based on the $do(x)$ operator.

3 The illusion of non-modularity

In her critics of the do -operator, Cartwright invokes yet another argument – the failure of modularity – which allegedly plague most mechanical and social systems.

In her words:

“When Pearl talked about this recently at LSE he illustrated this requirement with a Boolean input-output diagram for a circuit. In it, not only could the entire input for each variable be changed independently of that for each other, so too could each Boolean component of that input. But most arrangements we study are not like that. They are rather like a toaster or a carburetor.”

At this point, Cartwright provides a 4-equation model of a car carburetor and concludes:

“The gas in the chamber is the result of the pumped gas and the gas exiting the emulsion tube. How much each contributes is fixed by other factors: for the pumped gas both the amount of airflow and a parameter a , which is partly determined by the geometry of the chamber; and for the gas exiting the emulsion tube, by a parameter a' , which also depends on the geometry of the chamber. The point is this. In Pearl’s circuit-board, there is one distinct physical mechanism to underwrite each distinct causal connection. But that is incredibly wasteful of space and materials, which matters for the carburetor. One of the central tricks for an engineer in designing a carburetor is to ensure that one and the same physical design - for example, the design of the chamber - can underwrite or ensure a number of different causal connections that we need all at once.

Just look back at my diagrammatic equations, where we can see a large number of laws all of which depend on the same physical features - the geometry of the carburetor. So no one of these laws can be changed on its own. To change any one

requires a redesign of the carburetor, which will change the others in train. By design the different causal laws are harnessed together and cannot be changed singly. So modularity fails” [Cartwright, 2007, pp. 15-16].

Thus, for Cartwright, a set of equations that share parameters is inherently non-modular; changing one equation means modifying at least one of its parameters and, if this parameter appears in some other equation, it must change as well, in violation of modularity.

Heckman (2005, pp. 44) makes similar claims: “Putting a constraint on one equation places a restriction on the entire set of internal variables.” “Shutting down one equation might also affect the parameters of the other equations in the system and violate the requirements of parameter stability.”

Such fears and warnings are illusionary. Surgery, and the whole semantics and calculus built around it, does not assume that in the physical world we have the technology to incisively modify the mechanism behind each structural equation while leaving all others unaltered. Symbolic modularity does not assume physical modularity. Surgery is a symbolic operation that makes no claims about the physical means available to the experimenter, nor about possible connections that might exist between the mechanisms involved.

Symbolically, one can surely change one equation without altering others and proceed to define quantities that rest on such “atomic” changes. Whether the quantities defined in this manner correspond to changes that can be physically realized is a totally different question that can only be addressed once we have a formal description of the interventions available to us. More importantly, shutting down an equation does not necessarily mean meddling with its parameters; it means overruling that equation, namely, leaving the equation intact but lifting the outcome variable from its influence.

A simple example will illustrate this point.

Assume we have two objects under free fall condition. The respective accelerations, a_1 and a_2 of the two objects are given by the equations:

$$a_1 = g \tag{1}$$

$$a_2 = g \tag{2}$$

where g is the earth gravitational pull. The two equations share a parameter, g , and appear to be non-modular in Cartwright's sense; there is indeed no physical way of changing the gravitational force on one object without a corresponding change on the other. However, this does not mean that we cannot intervene on object 1 without touching object 2. Assume we grab object 1 and bring it to a stop. Mathematically, the intervention amounts to replacing Eq. (1) by

$$a_1 = 0 \tag{3}$$

while leaving Eq. (2) in tact. Setting g to zero in Eq. (1) is a symbolic surgery that does not alter g in the physical world but, rather, sets a_1 to 0 by bringing object 1 under the influence of a new force, f , emanating from our grabbing hand. Thus, Eq. (3) is a result of two forces:

$$a_1 = g + f/m_1 \tag{4}$$

where $f = -gm_1$, which is identical to (3).

This same operation can be applied to Cartwright's carburetor; for example, the gas outflow can be fixed without changing the chamber geometry by installing a flow regulator at the emulsion tube. It definitely applies to economic systems, where human agents are behind most of the equations; the left-hand side of the equations can be fixed by exposing agents to different information, rather than changing parameters in the physical world. A typical example emerges in job discrimination cases (ibid, p.128). To test the "effect of gender on hiring" one need not physically change applicant's gender; it is enough to change employers awareness of the applicant's gender.

This operation of adding a term to the right-hand side of an equation to ensure constancy of the left-hand side is precisely how Haavelmo (1943) envisioned surgery in economic settings. Why his wisdom disappeared from the teachings of his disciples in 2008 is one of the great mysteries of economics (see Hoover (2004)); my theory remains (ibid, p. 138) that it all happened due to a careless choice of notation which crumbled under the ruthless invasion of statistical thinking in the early 1970's. Still, I am yet to see an example of an economic system that is not modular in the sense described here.

4 Summary: Where is econometric modeling today?

By rejecting the surgical definition of structural counterfactuals Cartwright endangers econometrics with another decade of confusion and disputations.

In almost every one of his recent articles James Heckman stresses the importance of counterfactuals as a necessary component of economic analysis and the hallmark of econometric achievement in the past century. For example, the first paragraph of [Heckman and Vytlačil, 2007] reads: “they [policy comparisons] require that the economist construct counterfactuals. Counterfactuals are required to forecast the effects of policies that have been tried in one environment but are proposed to be applied in new environments and to forecast the effects of new policies.” Likewise, in his *Sociological Methodology* article (2005), Heckman states: “Economists since the time of Haavelmo (1943, 1944) have recognized the need for precise models to construct counterfactuals... The econometric framework is explicit about how counterfactuals are generated and how interventions are assigned...”

And yet, despite the proclaimed centrality of counterfactuals in economic analysis, a curious reader will be hard pressed to identify even one econometric article or textbook in the past 40 years in which counterfactuals or causal effects are formally defined. By rejecting Haavelmo’s surgery, Cartwright rejects what she calls “impostor counterfactuals” but fails to provide us with alternative definition of “genuine counterfactuals,” namely, a procedure for computing the counterfactual $Y(x, u)$ in a well-posed, fully specified economic model, with X and Y two arbitrary variables in the model.²

The absence of an explicit, formal definition for this fundamental quantity has allowed econometrics to split into two isolated, narrowly informed camps. Economists working within the Neyman-Rubin framework [Neyman 1923, Rubin 1974] take counterfactuals as primitive, unobservable variables, totally detached from the knowledge encoded in structural equation models (e.g., [Angrist, 2004, Imbens, 2004]). Even those working with propensity score techniques, whose validity rests entirely on the causal assumption of “ignorability,” or unconfoundedness, rarely know how to confirm (or invalidate) that assumption from structural

² $Y(x, u)$, sometimes written $Y_x(u)$, stands for the value that variable Y would attain in context u , had variable X been x . Pearl (2009) discusses Heckman and Vytlačil’s (2007) attempts to define this quantity using “external variations” and partial derivatives, instead of Haavelmo’s surgery.

knowledge. Economists working within the structural equations framework are busy estimating parameters while treating counterfactuals as metaphysical ghosts that should not concern ordinary mortals. They trust philosophers such as Cartwright and leaders such as Heckman to define precisely what the policy implications are of the structural parameters they labor to estimate, and to relate them to what their colleagues in the potential-outcome camp are doing.³

The surgery semantics (ibid, pp. 98-102) and the mathematical properties entailed by it, offer a simple and precise unification of these two estranged and poorly equipped schools of econometric research. Cartwright's (and Heckman's) objections to this semantics will not help these two schools realize that they are working on two aspects of the same mathematical object; a theorem in one framework is a theorem in another. Economists will do well resurrecting the basic ideas of Haavelmo (1943), Marschak (1950), and Strotz and Wold (1960) and re-invigorating them with the logic of graphs and counterfactuals developed in the past two decades.

References

[Angrist, 2004] J.D. Angrist. Treatment effect heterogeneity in theory and practice. *The Economic Journal*, 114:C52–C83, 2004.

[Cartwright, 2007] N. Cartwright. *Hunting Causes and Using Them: Approaches in Philosophy and Economics*. Cambridge University Press, New York, NY, 2007.

[Christ, 1966] C. Christ. *Econometric Models and Methods*. John Wiley and Sons, Inc., New York, 1966.

[Dawid, 2000] A.P. Dawid. Causal inference without counterfactuals (with comments and rejoinder). *Journal of the American Statistical Association*, 95(450):407–448, June 2000.

[Haavelmo, 1943] T. Haavelmo. The statistical implications of a system of simultaneous equations. *Econometrica*, 11:1–12, 1943. Reprinted in D.F. Hendry and M.S. Morgan (Eds.), *The Foundations of Econometric Analysis*, Cambridge University Press, 477–490, 1995.

³Anecdotically, the bibliographical list in the comprehensive review article by Hoover (2008) is almost disjoint from those of Angrist (2004) and Imbens (2004) – the cleavage is culturally deep.

- [Haavelmo, 1944] T. Haavelmo. The probability approach in econometrics (1944)*. Supplement to *Econometrica*, 12:12–17, 26–31, 33–39, 1944. Reprinted in D.F. Hendry and M.S. Morgan (Eds.), *The Foundations of Econometric Analysis*, Cambridge University Press, 440–453, 1995.
- [Heckman and Vytlacil, 2007] J.J. Heckman and E.J. Vytlacil. *Handbook of Econometrics*, volume 6B, chapter Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation, pages 4779–4874. Elsevier B.V., 2007.
- [Heckman, 2005] J.J. Heckman. The scientific model of causality. *Sociological Methodology*, 35:1–97, 2005.
- [Hoover, 2004] K.D. Hoover. Lost causes. *Journal of the History of Economic Thought*, 26(2), June 2004.
- [Hoover, 2008] K.D. Hoover. Causality in economics and econometrics. In Steven N. Durlauf and Lawrence E. Blume, editors, *From The New Palgrave Dictionary of Economics*. Palgrave Macmillan, New York, NY, 2nd edition, 2008.
- [Imbens, 2004] G.W. Imbens. Nonparametric estimation of average treatment effects under exogeneity: A review. *The Review of Economics and Statistics*, 86(1):4–29, 2004.
- [Leroy, 2002] S.F. Leroy. A review of Judea Pearl’s *Causality*. *Journal of Economic Methodology*, 9(1):100–103, 2002.
- [Marschak, 1950] J. Marschak. Statistical inference in economics. In T. Koopmans, editor, *Statistical Inference in Dynamic Economic Models*, pages 1–50. Wiley, New York, 1950. Cowles Commission for Research in Economics, Monograph 10.
- [Neyman, 1923] J. Neyman. On the application of probability theory to agricultural experiments. Essay on principles. Section 9. *Statistical Science*, 5(4):465–480, 1923.
- [Pearl, 2000a] J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, 2000.
- [Pearl, 2000b] J. Pearl. Comment on A.P. Dawid’s, causal inference without counterfactuals. *Journal of the American Statistical Association*, 95(450):428–431, June 2000.

[Pearl, 2003] J. Pearl. Reply to Woodward. *Economics and Philosophy*, 19:341–344, 2003.

[Pearl, 2009] J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, second edition, Forthcoming, 2009.

[Rubin, 1974] D.B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66:688–701, 1974.

[Strotz and Wold, 1960] R.H. Strotz and H.O.A. Wold. Recursive versus nonrecursive systems: An attempt at synthesis. *Econometrica*, 28:417–427, 1960.