ABSTRACT

There are simple mechanical systems that elude causal representation. We describe one that cannot be represented in a single directed acyclic graph. Our case suggests limitations on the use of causal graphs for causal inference and makes salient the point that causal relations among variables depend upon details of causal set ups, including values of variables.

Keywords: Causation, Causal Modeling, Counterfactuals

Systems without a Graphical Causal Representation

Daniel M. Hausman, Reuben Stern, Naftali Weinberger

According to Hausman (1998: 25-26)[1], causal relations among variables may differ depending on features of the specific systems in which they are instantiated. For example, in simple causal set ups involving gas cylinders subject to external determinants of their temperature or volume, the causal relations between pressure, temperature, and volume may differ. While an increase in temperature will lead to an increase in volume in a cylinder with a moveable piston, an increase in temperature will lead to an increase in pressure in a cylinder of fixed volume   Until one locates the variables, P, V, and T, within some system with a definite causal structure, the laws governing gases, such as the ideal gas law (PV = kT) tell one nothing about the direction of the causal relations.

In this essay, we expand on Hausman's thesis by showing that it implies that some systems cannot be represented in a single causal graph.[2] Although we do not know of any explicit claims to the contrary—i.e., that all causal systems can be represented by acyclic or cyclic graphs—we think that the example in this paper will come as a surprise to many working

---

[1] See also Woodward (2003: 234).
[2] This possibility is implied by an example Glymour considers (2010: 202).

in this field. Consider the device in figure 1 (call it "the elusive cylinder.") It consists of a cylinder of gas immersed in a water bath that is maintained at a constant temperature H. There is a piston at the top of the cylinder that can be locked into one of three positions (X = 1, 2, or 3) or allowed to move up or down depending on the pressure of the gas (X = 0) and on the weight placed on top of the piston. We assume that the piston moves up and down without friction and achieves a perfect seal and that the ideal gas law governs the relations among temperature, pressure, and volume.

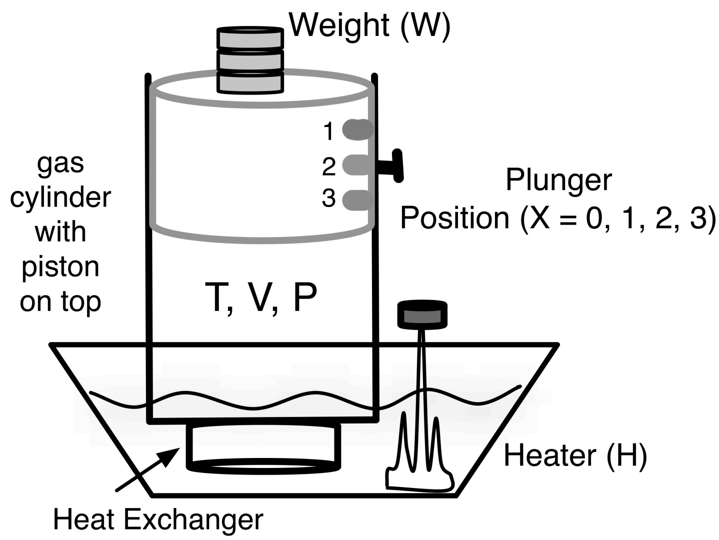

Figure 1: The Elusive Cylinder

If X ≠ 0 (that is, if the piston is locked into one of the three places), then the following directed acyclic graph correctly depicts how the cylinder will behave across interventions:[3]

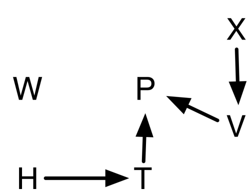

Figure 2: Locked Piston Graph (X≠0)

[3] Here we are interpreting directed acyclic graphs in the way specified by Pearl (2009) and Spirtes, Glymour, and Scheines (1993).

If instead X = 0 and the piston is not locked in place, then the causal relations differ and are represented by a different directed acyclic graph:

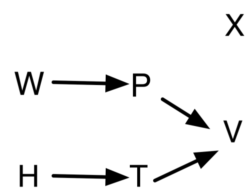

Figure 3: Floating Piston Graph (X=0)

Without specifying whether X = 0, the system has no representation as a directed acyclic graph: The variables other than H stand in no single causal relation to one another. Notice that this system does not illustrate the familiar point that enlarging causal graphs to include additional variables may lead one to correct mistakes. (For example, adding the variable C, which is a common cause of two variables, A and B that are not otherwise causally connected, permits one to delete a mistaken edge between A and B.) In the graphs of the elusive cylinder, in contrast, there is no misrepresentation owing to omitted variables: the direction of the causal arrow between P and V depends on the value of X. This dependence goes beyond so-called *effect modification* (VanderWeele and Robins 2007). Effect modification occurs when the magnitude of the effect of one variable on another depends on the value of a third variable. For example, whether flipping a switch causes a light to go on depends on whether there is power in the house. But, unlike the case of the elusive cylinder, in cases of effect modification causal arrows never change direction, and one can represent effect modifications by means of a directed acyclic graph.

Readers may wonder why, in principle, a DAG can represent other types of effect modification, but not our case. One cannot, of course, read off the functional relationships in a DAG from the graph alone. The graph tells one nothing about whether the relation between cause and effect is linear or whether there are causal interactions among the variables with edges into some effect. So the fact that the coefficient of the equation relating the light's being on to the switch position depends on the values of other variables does not rule out the possibility of representing these relations in a DAG. But no DAG can represent the causal relations in both the cases of the locked and floating pistons, because edges go in different directions in the two cases. If there is an arrow between P and V, but no corresponding arrow from V to P,[4] one cannot change V into a cause of P by intervening on either of the variables. An intervention on P would not alter the relationship at all. An intervention on V could change the relationship, but would not create a new asymmetric causal relationship from V to P.

While we have described the elusive cylinder as a single system that has no graphical representation, one could alternatively consider the locked piston and the floating piston graphs as representing two distinct systems. Accordingly, the elusive cylinder reveals that a simple mechanical set up can correspond to two distinct systems and that, surprisingly, it is possible to switch from one causal system to another without introducing any new variables or varying the background conditions. Under this description, each system has a causal representation, but there is no way to use a DAG to represent the effect of variation in the value of X on which system is instantiated. It follows that DAGs cannot represent all causal relationships, as they cannot capture the effects of X. Notice that the value of X need not be determined by an intervention. Its value could, for example, change periodically through some mechanism attached to a timer. We have no quarrel with defining "system" in such a way that all systems have some sort of graphical representation, but insofar as one can switch between systems by a change in the value of a variable, whether by intervention or via the operation of parts of the larger system, one ought to be able to represent the effects of this change.

[4] Below we consider the possibility that there are arrows going in both directions between P and V.

What's the upshot? First, the example underlines Hausman's original point: Only with reference to some specific set up (and in this case the value of the variable X) do their instantiations bear causal relations to one another.

Second, neither of the causal graphs enables one to model counterfactuals correctly concerning what would be the result of interventions on X. If initially the value of X is 2, one can use the locked piston graph to answer the question, "What if the value of X were set by intervention at 1 (or 3)?" but one must use the floating piston graph to answer the question, "What if the value of X were set by intervention at 0?" Nothing in either representation tells us which is the correct representation to use for which values of X. In Jim Woodward's language (2003), the causal relationships depicted in either graph are invariant to some interventions, but not to others. This fact might not seem disturbing: after all Hooke's Law does not tell us what will happen when one stretches a spring so far that it is deformed or broken. But in the failures of invariance that Woodward considers, the interventions destroy or modify a causal mechanism. Once one has stretched a spring too far, its elasticity changes; and when one stretches it beyond its breaking point, the spring ceases to obey Hooke's law altogether. In the elusive cylinder, by contrast, no causal mechanism is disrupted— although certain causal relations cease to obtain for some settings of X, one can restore these relations by changing the value of X.

Notice that the elusive cylinder differs from cases in which philosophers use separate graphs to distinguish systems in which one intervenes on a variable from systems without interventions. In those cases, there are theorems specifying the relationship between these graphs and there are "combined graphs" (Spirtes, Glymour and Scheines 1993: 77-78) of which the separate graphs are subgraphs. In our case, in contrast, there seems to be no formal way to determine how the variables in the graph will respond to interventions.

Third, this system poses problems of causal inference. If one has a set of measurements of all the variables (that is, of P, V, T, W, and X) and seeks to determine the causal relations among the variables from the covariance matrix, either one will be led to one or the other graph (if the value of X is rarely zero or rarely not zero), or one will be unable to determine the causal relations.

The bottom line is that the causal representation of the system by means of a DAG depends on the value of one of its variables. Without specifying the value of X, asymmetric causal relations in the elusive system are undetermined. Systems such as the elusive one either are not in some sense *causal* systems, or one needs some other mode of representation than the directed acyclic graph.

It is far from obvious what other mode of representation one can employ. Suppose, for example, one adopted the convention of drawing an arrow from a variable V to another variable U if and only if, for some value of a variable influencing the system (whether latent or manifest), V causes U. Adopting this convention yields the following representation of this system:

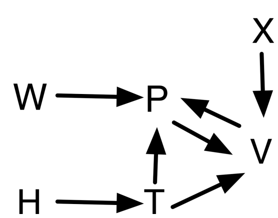

Figure 4: The Mixed Graph

The double arrow between P and V does not, of course, mean that there is any value of X such that it is both the case that P causes V and V causes P. For any way that the system is actually working, there is an asymmetric causal relation between P and V. It rather means that the direction of the causal arrow depends on the value of X. Similarly, for some values of X, some of the causal arrows (like the arrow from T to P) are “blocked” or inactive, while for other values, they are unblocked or active.

The mixed graph does nothing to address our concerns about grounding counterfactuals and predicting the outcomes of interventions. Suppose the piston is initially free to move, in which case X = 0. According to the mixed graph, the arrows from V to P, from T to P and from X to V (although present) are inactive. When one unlocks the piston (i.e., sets X = 1, 2 or 3), several changes occur: The three inactive arrows are all activated, and the active arrows from W to P, P to V, and from T to V are inactivated. It is not mysterious how setting X to 1, 2, or 3 could render the arrows from P to V or from T to V inactive. But setting the value of X = 1, 2, or 3 also inactivates the arrow from W to P and activates the arrow from V to P. These ramifications are far beyond those ordinarily attributed to interventions, whether interpreted as arrow-breaking or as mere additional causal influences.[5] Nothing in the mixed graph justifies predicting these consequences.

Conversely, suppose the piston is initially sealed in one of the three positions. It would seem that the arrows from V to P and T to P would be active while the arrows from P to V and W to P would be inactive. Then, were we to intervene on the system to set X to 0, it would seem that the arrows from V to P and T to P would be deactivated while the arrow from W to P would be activated. But, again, these are not the ordinary predictable consequences of interventions. An intervention that deactivates a cause of V cannot modify the effects of V, T and W on P without influencing a cause of P, and X is not a cause of P. We conclude, then, that the change that results from locking or unlocking the piston cannot be represented as an intervention. The mixed graph sheds no light on how changes in X influence the other variables in the system.

[5] Korb et al. (2004) model interventions in such a way that they do not necessarily wipe out a variable's relationship to its prior causes, but might merely alter its probability distribution. But even interventions of this sort are incapable of rendering some of a variable's causes inactive while activating others. See also Eberhardt and Scheines (2007).

To forestall confusion, we should emphasize that we are not denying that there are cases in which one would properly represent a system with a graph that contains cycles. Even if every token causal relation in a system is asymmetric, there will often be a feedback loop between two variables such that in the graph representing the type-level relations between the variables there will be an arrow from each of the variables to the other. For example, education at a time will increase income at a later time, which in turn will increase the amount of education one pursues later on. If the variables in one's graph were just education and income, there would be an arrow from each variable to the other. Although one often uses DAGs for the sake of simplicity, one can use cyclic graphs as well.[6] We find the mixed graph unsatisfying not because of a prejudice against cycles, but rather because there is no cycle in the case of the elusive cyllinder. For any value of X, the arrow between P and V is asymmetric.

In light of the difficulties with representing the elusive cylinder, one might respond that it is not a causal system and that it is unsurprising that no DAG can represent it. But setting X determines the causal relations, and some formal encoding of the consequences of setting X is needed. Current graphical representations of causal relations are inadequate to the task. The elusive cylinder eludes causal representation.

References

Eberhardt, F., & Scheines, R. (2007). Interventions and causal inference. *Philosophy of Science*, *74*(5), 981-995.

Glymour, Clark, Richard Scheines, and Peter Spirtes. (2001) *Causation, Prediction, and Search*. MIT Press.

Glymour, Clark. (2010). What Is Right with 'Bayes Net Methods' and What Is Wrong with

[6] See, for example, Strotz and Wold (1960) and Glymour, Scheines and Spirtes (2001, 297-9).

‘Hunting Causes and Using Them’? *British Journal for Philosophy of Science* 61**:** 161–211.

Hausman, Daniel. (1998). *Causal Asymmetries.* Cambridge: Cambridge University Press.

Hausman, Daniel. (2005). Causal Relata: Tokens, Types or Variables. *Erkenntnis* 63: 31-51.

Pearl, Judea. (2009). *Causality: Models, Reasoning, and Inference.* 2nd ed. Cambridge: Cambridge University Press.

Korb, K. B., Hope, L. R., Nicholson, A. E., & Axnick, K. (2004). Varieties of causal intervention. In *PRICAI 2004: Trends in Artificial Intelligence* (pp. 322-331). Springer Berlin Heidelberg.

Spirtes, Peter, Clark Glymour, and Richard Scheines. (1993). Causation, Prediction and Search. New York: Springer-Verlag.

Strotz , R. H. , and H. O. A. Wold. (1960). Recursive versus nonrecursive systems: An attempt at synthesis. *Econometrica* 28: 417-27

VanderWeele, Tyler J., and James M. Robins. (2007). Four types of effect modification: A classification based on directed acyclic graphs. *Epidemiology* 18.5: 561-568.

Woodward, James. (2003). *Making Things Happen.* Oxford: Oxford University Press.