Race Causality Discrimination – Manipulation (cont'ed)

Marcello Di Bello - ASU - Fall 2023 - Week #3

This week we continue to examine the manipulationist account of causation. Last week we looked at Rubin model of causation as described by statistician Paul Holland[1]. This week we turn to the philosophical literature, in particular, James Woodward's manipulationist account of causation.[2]

[1] Holland (1986), Statistics and Causal Inference, *Journal of the American Statistical Society*, 81(396).

[2] James Woodward (2003), *Making Things Happen: A Theory of Causal Explanation*, 2003, Oxford University Press. We'll focus on chapter 2.

The account in short

The basic idea of Woodward's account is this. For X to be a cause of Y, the following counterfactual should hold:

> If the value of X were to change as a result of *some*[3] intervention, then the value of Y would change, as well.

If there is no intervention such that changing the value of X would also change the value of Y, then X isn't a cause of Y.[4] Woodward is clear that his account of causation is not completely reductive, since the notion of manipulation is itself a causal notion.[5]

[3] The fact that causal claims are valid only under some, and not necessarily all interventions, is significant. How so? See section 2.5.

[4] Crucially, this is not a merely counterfactual account of causality. The requirement for X to be a cause of Y is not simply that, if the value of X were to change, then the value of Y would change as well. What is the difference? Another point to note here is that Woodward is primary offering an accounting of type-level causation. For his account of token-level (or actual) causation—which presupposes his account of type causation, see section 2.7.

[5] How does Woodward address the cnharge of circularity, then? See section 1.7, especially p. 22.

Graphs and Equations

Before we look at a few examples, it is useful to become familiar with the formal framework. The framework consists of graphs and equations. Graphs will consist of vertices (variables) and directed edges (arrows). An arrow from X to Y simply means that there is a causal relational from X to Y. Graph can have different, ever more complex structures. One vertex Z can have two incoming arrows, one from X and another from Y, like this:

$$X \rightarrow Z \leftarrow Y.$$

This would be a *common effect* scenario. One vertex X can have two outgoing arrows, like this:

$$Y \leftarrow X \rightarrow Z.$$

This would be a *common cause* scenario. Finally, two edges could be mediated by a third, like this:

$$X \rightarrow Z \rightarrow Y.$$

This would be a *causal chain* scenario.[6]

[6] As we look at some examples below, it is instructive to draw the appropriate causal graph to get an intuitive picture of what is going on.

Compared to causal graphs, equations provide a more fine-grained information about the causal relationships between variables. For consider the following two equations:

$$Z = aX + bY$$

$$Z = XY$$

They could both be represented by a common effect causal graph, but this would miss out crucial information. The first equation is additive: any change in X or Y result in a corresponding change in Z discounted by the coefficients a and b. The second equation is a product. Suppose Z, X and Y can take only value 1 (present) or 0 (absent). Then if both X and Y equal 1, then Z will also equal 1, but if either X or Y equal 0, then Z will also be 0 There is a particular interaction between X and Y that brings about Z.[7]

[7] For a realistic example of this, see equation (2.2.3) on p. 44.

Example 1: Barometers

> suppose that, in a certain region, changes in atmospheric pressure (A) are a common cause of the occurrence of storms (S) and of the reading (B) of a particular barometer (p. 46)

There is a counterfactual dependence between variables B and S, since changes in B correlate to changes in S. Had the barometer reported a different atmospheric pressure, the whether would have been different. But this does not mean that B counts as a cause of S, despite the counterfactual dependence between the two variables. How do we account for this example?

Intervention and direct causes

Woodward argues that no (properly defined) intervention that would change B would also change S. So, B does not cause S as expected. The trick here is to be clear about what counts as an *intervention* on B. Here is what does not count as a (proper intervention) on B:

> there are ways of changing B that will be associated with a corresponding change in S even though B does not cause S. For example, if we change B by changing A (p. 46).

So what does count as a proper (legitimate, acceptable) intervention on B? Here is Woodward's sketch of the answer:[8]

[8] A fuller answer is given in section 3.1, non assigned as a reading for this class.

> interventions involve exogenous changes in the variable intervened on. When an intervention occurs on B, the value of B is determined entirely by the intervention, in a way that is (causally and probabilistically) independent of the value of A. In this sense, the intervention "breaks" the previously existing endogenous causal relationship between A and B. More generally and slightly more precisely, we may think of an intervention on X with respect to Y as an exogenous causal process that changes X in such a way and under conditions such that if any change occurs in Y, it occurs only in virtue of Y's relationship to X and not in any other way. (p. 47)

So, more generally, an intervention on a variable X in order to change another variable Y should be exogenous to the existing causal relations that hold between X and Y. An intervention on X is a change of X that holds fixed all other variables, except X and Y. This leads to the following definition of direct cause (DC):

> (DC) A necessary and sufficient condition for X to be a direct cause of Y with respect to some variable set V is that there be a possible intervention on X that will change Y (or the probability distribution of Y) when all other variables in V besides X and Y are held fixed at some value by interventions.

Note that the (i) the intervention on X must hold fixed all other variables except X and Y and (ii) the variables are relative to a variable set V (more on this later).[9]

[9] The importance of the choice of a variable set is discussed extensively in section 2.8 on the topic of serious possibilities.

Example 2: Birth control and thrombosis

Consider this example:

> Suppose that birth control pills (B) directly cause an increased probability of thrombosis (T) but also directly lower the probability of pregnancy (P), which is itself a direct positive probabilistic cause of thrombosis. As it happens, the probability increase in T that is directly due to B is exactly balanced by the decrease in probability of T which occurs along the B->P->T route, so that the net change in the probability of thrombosis if one takes birth control pills is 0. (p. 49)

What's interesting about this example is that—*all things considered*—B does not cause T, and yet there is a sense in which B does cause T. Now, according to the above definition of direct cause, B is indeed a direct cause of T. If we intervene on B, while other all other variables are fixed (including pregnancy P), then indeed a change in B would result in a change in T. So, the sense in which B is a cause of T is captured by the notion of direct cause.

Contributing causes

Does the notion of direct cause suffice to capture causality as such? It does seem to be sufficient because a variable might cause another via a third intermediate variable (recall the chain: $X \to Z \to Y$). We can call this a *contributing cause*.

So for X to cause Y (necessary conditions), we require the following (p. 57):

- X is a direct cause of Y (as defined above); or
- there is causal chain between X and Y, each link of which involves a relation of direct causation.

Note that these conditions only offers a necessary conditions for causality to hold between X and Y. Do they also offer sufficient conditions? That is, whenever there is direct causation from X to Y or a causal chain of direct causal link, does it follow that X causes Y? Not necessarily, as we will now see.

Example 3: Bites and bombs

Consider this example:

> A dog bites off my right forefinger. The next day I detonate a bomb by using my left forefinger. If I had not lost my right finger, I would have used it instead to detonate the bomb. The bite causes me to use my left finger, which causes the bomb to explode, but (it seems) the bite does not cause the bomb to explode.

There is a direct causation between the bite B and the fact that I am using my left finger to detonate the bomb (F). There is also a direct causation between the fact that I am using my left finger to detonate the bomb and the explosion E. Yet, intuitively, the bite does not seem to have caused the explosion. This example challenges the transitivity of causality, but also shows that the existence of a causal chain of direct causal links between X and Y is not enough to establish causality between X and Y. What else is needed?

> for X to be a contributing cause of Y, not only must there be at least one chain of direct causal relationships (a directed path or route) from X to Y,but it must also be the case that the value of Y is sensitive along that path to some interventions that change the value of X (p. 59)[10]

[10] How does this account deal with the bomb example? A more precise statement of the account is given on p. 59 under M (standing for "manipulability theory"). Make sure you understand every single statement in M.

Example 4: far-fetched survival

An objection to the manipulability account is that it is relative to a set of variables of interest and thus subjective. Consider this example:

> X lives at a great distance from the patient, has no responsibility for his care, is not aware that he exists, and is not a doctor. Suppose that if X had happened to go the hospital room where the patient is being cared for and had realized that the patient was developing a fever and had learned that administration of the antibiotic was the appropriate response and had administered the drug, the patient would have survived: the patient's survival is counterfactually dependent on whether X does these things. Nonetheless, we are not ordinarily inclined to think that X's failure to do these things caused the patient's death. (p. 88)

Even though X and the patient's survival are counterfactually dependent, it is far-fetched to claim that X could cause the patient' survival. What us going on here? Are causal claims dependent on moral considerations, customs and a social expectations?[11]

[11] To save objectivity, Woodward writes (p. 90): "the structures of counterfactual dependence ... hold independently of which possibilities we are willing to take seriously; in this sense, they are interest-independent and objective ... By contrast, causal judgments reflect both objective patterns of counterfactual dependence and which possibilities are taken seriously"