

Causation and Manipulability

- Administrative: (i) please post for this week, (ii) volunteer for Thursday, (iii) “minutes” from last week online, (iv) Some new papers online (see readings for weeks 2 and 3). In particular, Woodward and Hitchcock on Causation and Explanation, and Menzies on contextuality.
- Two Manipulability Accounts of Causation
 - von Wright
 - Menzies and Price
- Two Problems for Manipulability Accounts
 - Circularity
 - Unmanipulable causes
- The Interventionist Account of Woodward & Hitchcock
- Comparison with Counterfactual Account (an example)

von Wright — The Basic Idea — Quotes

Idea: “... to think of a relation between events as causal is to think of it under the aspect of (possible) action. It is therefore true, but at the same time a little misleading to say that if p is a (sufficient) cause of q , then if I could produce p I could bring about q . For that p is the cause of q , I have endeavored to say here, means that I could bring about q , if I could do (so that) p .”

Non-Circularity: “The connection between an action and its result is intrinsic, logical and not causal (extrinsic). If the result does not materialize, the action simply has not been performed. The result is an essential “part” of the action. It is a bad mistake to think of the act(ion) itself as a cause of its result.”

Unmanipulables: “The eruption of Vesuvius was the cause of the destruction of Pompeii. Man can ... destroy cities, but he cannot, we think, make volcanoes erupt. Does this not prove that the cause-factor is not distinguished from the effect-factor by being in a certain sense capable of manipulation? The answer is negative. The eruption of a volcano and the destruction of a city are two very complex events. Within each of them a number of events or phases and causal connections between them may be distinguished ... such that the cause-factor typically satisfies the requirement of manipulability.”

von Wright — The Basic Idea — Critical Summary I

- The first obvious charge against this view is that it is viciously circular, since the notion of “bringing about” is already causal notions.
- von Wright responds to this by saying that “action a brings about event b ” or “ b is the result of action a ” is to be understood as a *logical* relation between an action and its result. He suggests that the result of an act is an essential “part” of the act itself. So, it’s a category mistake to think of results of acts as (causal) *effects* of acts.
- How do we identify “the results” of actions? Do we just go by an agent’s descriptions of actions? May this include unintended consequences of actions? It seems not, since it’s hard to see how an unintended consequence of an action could be considered an “essential part” of it.
- Also, this is a deterministic account, since, according to vW, “If the result does not materialize, the action simply has not been performed.”

von Wright — The Basic Idea — Critical Summary II

- The second problem here involves unmanipulable causes. Intuitively, there will be causes which are not manipulable.
- von Wright tries to avoid this by reducing such cases (*e.g.*, galaxies attracting one another gravitationally) to “similar” cases (*e.g.*, people and projectiles falling to earth) in which the causes are manipulable.
- As Woodward points out, this leads to a dilemma:
 “Either we explain the relevant notion of similarity in straightforwardly causal terms that seem to have nothing to do with agency (*e.g.*, we say that the similarity consists in the fact that the same gravitational force law is operative in both cases), in which case we have effectively abandoned the agency theory, or else we are led to the conclusion that causal claims involving unmanipulable causes like galaxies involve a conception of causality which is fundamentally different from the conception that is applicable to manipulable causes.”

Menzies and Price — Refinements? — Quotes

Idea #1: “... an event A is a cause of a distinct event B just in case bringing about the occurrence of A would be an effective means by which a free agent could bring about the occurrence of B .”

Idea #2: “[a]gent probabilities are to be thought of as conditional probabilities, assessed from the agent’s perspective under the supposition that antecedent condition is realized *ab initio*, as a free act of the agent concerned. Thus the agent probability that one should ascribe to B conditional on A is the probability that B would hold were one to choose to realize A .”

Non-Circularity: “... we have direct experience not merely of the Humean succession of events in the external world, but of a very special class of such successions: those in which the earlier event is an action of our own, performed in circumstances in which we both desire the later event, and believe that it is more probable given the act in question than it would be otherwise. ... [this] does not depend on prior acquisition of any causal notion. An agency theory thus escapes the threat of circularity.”

Unmanipulables “... we would argue that when an agent can bring about one event as a means to bringing about another, this is true in virtue of certain basic intrinsic features of the situation involved, these features being essentially non-causal though not necessarily physical in character. Accordingly, when we are presented with another situation involving a pair of events which resembles the given situation with respect to its intrinsic features, we infer that the pair of events are causally related even though they may not be manipulable. ... events such as the claim that the 1989 San Francisco earthquake was caused by friction between continental plates. We can make such causal claims because we believe that there is another situation that models the circumstances surrounding the earthquake in the essential respects ... The paradigm example of such a situation would be that created by seismologists in their artificial simulations of the movement of continental plates.

Menzies and Price — Refinements? — Critical Summary I

- M&P give an account of the following form:
 a causes b iff $\Pr(b \mid do(a)) > \Pr(b \mid \neg do(a))$.
- Here, $\Pr(b \mid do(a))$ is the probability a free (rational?) agent assigns to the result b obtaining, conditional on the agent’s choosing to realize a .
- This now makes the theory of causation depend on an (rational?) agent’s degrees of belief about (their) actions and their consequences.
- One advantage over vW is that the theory of M&P is consistent with *indeterministic* causation. But, the theory seems more a theory of *causal belief* than a metaphysical theory about the nature of causation.
- How are we to interpret $\Pr(b \mid do(a))$, precisely? Is this the degree of belief in b on the *supposition* of $do(a)$? Or, is it the degree of belief in b one would have if one were to *learn* that $do(a)$ was actually the case?
- What work is the “freedom” of the agent doing here?

Menzies and Price — Refinements? — Critical Summary II

- M&P handle the circularity objection a bit differently than vW does. They argue that we have an understanding of what $\Pr(b \mid do(a)) > \Pr(b \mid \neg do(a))$ means, *independently* of any causal notions.
- Both vW and M&P are committed to the claim that the notions of “agency” and “manipulability” are at least as primitive (if not more so) than the notion of causation. Is this correct?
- M&P have an interesting response to the unmanipulables problem. They also try to reduce cases of unmanipulable causation to “similar” cases which are manipulable. They use simulations as their guiding example. But, simulations are not perfect, and we need a reason to think that they will capture the (intuitive) causal structure of the real systems.
- M&P give no reason to believe that the requisite resemblance can be given a non-causal accounting, and that the resemblance relations can be understood solely on the basis of what we know about manipulation.

Counterfactuals to the Rescue?

- Perhaps we can avoid the problem of unmanipulables by making the account counterfactual. We could try: A causes B if and only if B would change if an appropriate manipulation on A were to be carried out.
- This is consistent with human beings not being able to manipulate, or even with no human beings existing. So far, so good.
- But, there are serious problems with trying to understand and assess such counterfactuals. What sort of situation or possibility we are supposed to envision when we imagine that the antecedent is true?
- Woodward asks us to consider “a causal claim about the very early universe during which temperatures are so high that atoms and molecules and presumably anything we can recognize as an agent cannot exist. What counterfactual scenario or possible world are we supposed to envision when we ask what would happen if human beings were to exist and were able to carry out certain manipulations in this situation?”

Spurious Correlations and M&P's Account I

- Let atmospheric pressure be represented by a (continuous) variable Z , the reading of a barometer be represented by a (continuous) variable X , and the occurrence of a storm be represented by a (dichotomous) variable Y , with no causal relationship between X and Y .
- X and Y will be correlated, but M&P's suggestion is that, conditional on the realization of X by a free act [$do(X = x)$], this correlation will disappear, indicating that the correlation between X and Y is spurious and does not reflect a causal connection from X to Y .
- If, by contrast, this correlation were to persist conditional on $do(X = x)$, this would be an indication that X was after all a cause of Y .
- That is, M&P claim we will have $\Pr(Y | X = x) > \Pr(Y)$. But, if X is not a cause of Y , then M&P claim we will have $\Pr(Y | do(X = x)) = \Pr(Y)$.
- How is this ensured by M&P's “free action” account? It isn't ...

Spurious Correlations and M&P's Account II

- “Free” usually means (i) uncoerced or unconstrained or due to voluntary choices of the agent, or (ii) uncaused or not deterministically caused.
- Neither of these is sufficient to avoid spurious correlation. Suppose that X is set by a free act (in either of the above senses) but that this free act (and hence X) is correlated with Z (perhaps because the experimenter observes Z and freely chooses to set X in a way that is correlated with Z — this could still be “free” in either sense). In this case, $\Pr(Y | do(X = x)) > \Pr(Y)$, even though X does not cause Y .
- Reply: require that $do(A)$ not be correlated with any other cause of A . Insufficient: $do(A)$ which produces A can also cause B via a route that does not go through A . Consider a case in which an experimenter's administration of a drug to a treatment group has a placebo effect that enhances recovery, even though the drug itself has no effect on recovery. Here, $\Pr(B | do(A)) > \Pr(B)$, even though A does not cause B .

The Interventionist Account: Woodward & Hitchcock I

- W&H put the following four constraints on an *intervention* I (i.e., $do(X = x)$), to try to handle all cases of spurious correlation:
 - M1. I must be the only cause of X ; i.e., the intervention must completely disrupt the causal relationship between X and its previous causes so that the value of X is set *entirely* by I ,
 - M2. I must not directly cause Y via a route that does not go through X as in the placebo example,
 - M3. I should not itself be caused by any cause that affects Y via a route that does not go through X , and
 - M4. I leaves the values taken by any causes of Y except those that are on the directed path from I to X to Y (should this exist) unchanged.
- The notion of a *causal effect* is defined in terms of the *difference* made to the value of Y by a change or difference in the value of X .
- Focusing on differences in this way allows us to isolate the contribution made to Y by X from the contribution made to Y by its other causes.

The Interventionist Account: Woodward & Hitchcock II

- Since in the non-linear case, the change in the value of Y caused by a given change in the value of X will depend on the values of the other causes of Y , the notion of causal effect must be *relativized to a background context* B which incorporates information about these other values.
- In deterministic contexts, the causal effect on Y of a change in the value of X from x to x' in circumstances B is: $Y_{do(X=x) \& B} - Y_{do(X=x') \& B}$ — the difference between the value that Y would take under $do(X = x)$ in circumstances B and the value that Y would take under $do(X = x')$ in B , where interventions $do(\mathcal{X} = \chi)$ are understood as in (M1) – (M4).
- In non-deterministic contexts, the definition of causal effect is, analogously, $Pr(Y = y \mid do(X = x) \& B) - Pr(Y = y \mid do(X = x') \& B)$.
- X is a cause of Y in B iff there are distinct values of X , x and x' such that the value of (probability distribution of) Y under $do(X = x)$ in B is different from the value of (prob. distribution of) Y under $do(X = x')$.

Brief Comparison with Lewisian Counterfactual Accounts: EPR

- In EPR type set-ups, as a matter of physical law alone there is no intervention on the outcome of the left hand measurement, with respect to the outcome of the right hand measurement.
- Any method of bringing about a desired outcome on the left hand side (such as preparing the particle pairs in a state other than the singlet state) will have a direct effect on the outcome on the right hand side too.
- It is thus natural to think of this as a case in which the perfect (anti)-correlation between the two measurement outcomes in an EPR experiment is not invariant under (any) interventions, and it follows that one measurement result does not cause or explain the other.
- Contrast this with the consequence of Lewis's account that the outcome on one wing causes the outcome on the other — (perfect) correlations in quantum mechanics are counterfactual supporting. See Jeremy Butterfield's 1992 paper "David Lewis Meets John Bell" for discussion.