Philosophy of causality

Douglas Ezra Morrison

Contents

1 Why look at philosophy at all? 1

2 Five accounts of causation 2
- 2.1 Regularity . . . 2
- 2.2 Counterfactual dependence . . . 2
- 2.3 Manipulation and intervention . . . 3
- 2.4 Probabilistic causation . . . 3
- 2.5 Mechanisms . . . 4

3 What the potential-outcomes framework takes from these accounts 4

4 Not everyone agrees 5

5 Further reading 6

> **Optional reading**
>
> This chapter is background. Nothing later in the book depends on it. It explains where the potential-outcomes definitions in the introduction[a] come from, and why they carry the assumptions they do.
>
> [a] intro-causal.qmd

1 Why look at philosophy at all?

The book defines a cause in terms of potential outcomes (the definition of a cause[1]). That definition is one answer among several to a question philosophers have argued over for centuries: what does it mean to say that one thing causes another? Seeing the other answers makes it easier to see what the potential-outcomes answer includes, what it leaves out, and why.

Each account below is a family of views, and this chapter simplifies all of them. The cited sources give the full arguments.

[1] intro-causal.qmd#def-cause

2 Five accounts of causation

2.1 Regularity

Definition 2.1 (Regularity view of causation)**.** On the **regularity view**, saying that A causes B means that every event like A is followed by an event like B (a "constant conjunction"), with A coming first and the two occurring close together in space and time.

Hume's discussion of causation is the standard starting point (Hume 1748, sec. 7). His point is that we never observe a necessary connection between cause and effect. We observe only that one kind of event reliably follows another.

Example 2.1.

2.1.1 Regularity and smoking

Under a strict reading of the regularity view, "smoking causes lung cancer" would be false, because many smokers never develop lung cancer and some non-smokers do.

Example 2.1 shows the main weakness: strict regularities are rare in epidemiology, where outcomes are stochastic and have many contributing causes. Mackie's refinement treats a cause as an insufficient but necessary part of a condition that is itself unnecessary but sufficient for the effect (Mackie 1965). That version fits "component causes" better.

2.2 Counterfactual dependence

Definition 2.2 (Counterfactual account of causation)**.** On the **counterfactual account**, event A causes event B if A and B both occurred, and, had A not occurred, B would not have occurred.

Hume also wrote a second, counterfactual formulation of his definition (Hume 1748, sec. 7). Lewis later developed it into a full theory based on comparing the actual world with the closest possible worlds in which A does not occur (Lewis 1973).

Example 2.2.

2.2.1 Counterfactual account and smoking cessation

Suppose a patient quit smoking and did not develop lung cancer. On the counterfactual account, quitting prevented the cancer only if the patient would have developed lung cancer had they kept smoking. We never observe that second scenario. It is the counterfactual outcome[2].

[2] intro-causal.qmd#def-counterfactual

This account is the closest to the definition of a cause[3]. A potential outcome $Y(a')$ for an action a' that did not occur is a counterfactual outcome, so the potential-outcomes framework makes counterfactual dependence something a statistical model can describe (Holland 1986).

The account has two well-known difficulties. First, "the closest possible world" is a philosophical notion that has no obvious empirical meaning. Second, it handles redundant causes badly: if two sufficient causes were both present, removing either one leaves the effect in place, so neither counts as a cause. The potential-outcomes framework avoids the first difficulty by defining the counterfactual in terms of a specific action (actions[4], Section 2.3), and mostly sidesteps the second by asking about the effect of one well-specified action at a time.

2.3 Manipulation and intervention

Definition 2.3 (Manipulability account of causation)**.** On the **manipulability account**, A is a cause of B if an intervention that changes A, and that affects nothing else that bears on B, would change B (Woodward 2003).

Here the cause is something that someone could in principle change. Holland summarized this view as "no causation without manipulation" (Holland 1986).

Example 2.3.

2.3.1 Manipulability and a patient's sex

"What is the effect of a smoking-cessation program on lung cancer risk?" names an intervention that can be assigned, so it fits the manipulability account.

"What is the effect of a patient's sex on lung cancer risk?" does not name an intervention. On this account the question is not well posed until we say which intervention is meant.

This account is the one that the book's definitions (actions[5] and causes[6]) take on: a cause is an *action*, and the effect is a contrast between outcomes under different actions.

2.4 Probabilistic causation

Definition 2.4 (Probabilistic view of causation)**.** On the **probabilistic view**, A causes B if A raises the probability of B after other relevant factors are held fixed (Suppes 1970).

This view repairs the regularity view for stochastic outcomes. It also matches how epidemiologists work: a risk factor changes the probability of disease and does not determine it.

Example 2.4.

[3] intro-causal.qmd#def-cause
[4] intro-causal.qmd#def-action
[5] intro-causal.qmd#def-action
[6] intro-causal.qmd#def-cause

2.4.1 Probability raising and smoking

Smoking raises the probability of lung cancer among people who are alike in age and other relevant factors, and it does not make lung cancer certain. Under the probabilistic view, that is enough for smoking to be a cause of lung cancer.

Its main difficulty is the one every observational study faces. Probability raising can come from confounding, so "after other relevant factors are held fixed" does the heavy lifting, and the view does not say which factors those are. The chapter on bias analyses[7] discusses how to handle that question in practice.

2.5 Mechanisms

Definition 2.5 (Mechanistic view of causation)**.** On the **mechanistic view**, A causes B if there is a mechanism, an organized set of entities and activities, that links A to B (Machamer et al. 2000).

Example 2.5.

2.5.1 Mechanism and smoking

A mechanistic account of smoking and lung cancer would describe the steps between inhaling smoke and the growth of a tumor, such as damage to cells' DNA. Evidence about such steps can support the causal claim alongside the statistical evidence.

Mechanistic evidence, such as a known biological pathway, often strengthens a causal conclusion even when the statistical evidence is limited. The potential-outcomes framework does not require a mechanism to define an effect. It defines the effect of an action on an outcome without describing how the action works.

3 What the potential-outcomes framework takes from these accounts

Table 1 summarizes how each account relates to this book.

Table 1: Philosophical accounts of causation and where they appear in this book.

Account	Core idea	Relation to this book
Regularity	Events of type A are followed by events of type B	Motivates asking whether an association is causal; too strict for stochastic outcomes

[7] bias-analyses.qmd

Account	Core idea	Relation to this book
Counterfactual	B would not have occurred without A	Basis of potential outcomes[8] and counterfactual outcomes[9]
Manipulability	A cause is something one could change	Basis of the definitions of actions[10] and causes[11]
Probabilistic	Causes raise probabilities	Matches estimands defined on risks and means
Mechanistic	Causes work through mechanisms	Not needed for definitions; useful as supporting evidence

Two assumptions in the book come straight from the counterfactual and manipulability accounts:

- **Consistency** (definition[12]) says that the outcome we observe under action a equals the potential outcome $Y(a)$. That statement is only meaningful if "action a" names one thing. Hernán and Taubman (2008) and Hernán (2016) argue this using examples such as obesity and body mass index: different ways of changing the exposure can lead to different outcomes, so a question about "the" effect of the exposure is ambiguous until the intervention is specified. VanderWeele and Hernán (2013) formalize the point as the problem of multiple versions of treatment.
- **Well-defined interventions** are the practical form of the manipulability account. If we cannot say what intervention would set A to a, we cannot say what $Y(a)$ refers to, and consistency has nothing to hold on to.

Example 3.1.

3.0.1 A well-defined and an ill-defined question

"What would a patient's blood pressure be if they took 10 mg of a particular antihypertensive daily for a year?" names a specific intervention, so $Y(a)$ is meaningful.

"What would a patient's blood pressure be if they had a lower body weight?" does not. Diet, exercise, surgery, and illness all lower weight, and they plausibly affect blood pressure differently. We would need to specify a particular intervention to define the causal question.

4 Not everyone agrees

The manipulability requirement is contested. Some authors hold that causal questions about attributes such as race or sex are meaningful without a specified intervention. Others reply that such questions can only mean something once they are tied to an intervention on a related, manipulable variable. This book follows the potential-outcomes tradition and

[8] intro-causal.qmd#def-potential-outcome
[9] intro-causal.qmd#def-counterfactual
[10] intro-causal.qmd#def-action
[11] intro-causal.qmd#def-cause
[12] intro-causal.qmd#def-consistency

requires a specified intervention. That choice is a design decision for the course, and the other accounts above remain legitimate frameworks.

5 Further reading

- The original discussion of the regularity and counterfactual formulations is Hume's *Enquiry* (Hume 1748).
- Lewis (1973) is the standard statement of the counterfactual account.
- Woodward (2003) develops the interventionist (manipulability) account in detail.
- Holland (1986) connects the philosophical discussion to statistical causal inference.
- Hernán (2016) gives an epidemiologist's argument for well-defined interventions.

Hernán, M. A., and S. L. Taubman. 2008. "Does Obesity Shorten Life? The Importance of Well-Defined Interventions to Answer Causal Questions." *International Journal of Obesity* 32 (S3): S8–14. https://doi.org/10.1038/ijo.2008.82.

Hernán, Miguel A. 2016. "Does Water Kill? A Call for Less Casual Causal Inferences." *Annals of Epidemiology* 26 (10): 674–80. https://doi.org/10.1016/j.annepidem.2016.08.016.

Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81 (396): 945–60. https://doi.org/10.1080/01621459.1986.10478354.

Hume, David. 1748. *An Enquiry Concerning Human Understanding*. London.

Lewis, David. 1973. "Causation." *The Journal of Philosophy* 70 (17): 556–67. https://doi.org/10.2307/2025310.

Machamer, Peter, Lindley Darden, and Carl F. Craver. 2000. "Thinking about Mechanisms." *Philosophy of Science* 67 (1): 1–25. https://doi.org/10.1086/392759.

Mackie, J. L. 1965. "Causes and Conditions." *American Philosophical Quarterly* 2 (4): 245–64.

Suppes, Patrick. 1970. *A Probabilistic Theory of Causality*. North-Holland.

VanderWeele, Tyler J., and Miguel A. Hernán. 2013. "Causal Inference Under Multiple Versions of Treatment." *Journal of Causal Inference* 1 (1): 1–20. https://doi.org/10.1515/jci-2012-0002.

Woodward, James. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press.