

Causes, Interventions, and Responsibility

Enno Fischer

Causes, Norms, and Decisions – Hannover – 15.8.2018



DFG-Graduiertenkolleg GRK 2073
Integrating Ethics and Epistemology of Scientific Research



Aims

1. Introduce a new taxonomy of actual causation.
2. Point out the limits of purely interventionist accounts of the function of causal reasoning.
3. Show that including responsibility helps where purely interventionist accounts reach their limits.

- 1 The Interventionist Theory
- 2 Criticism
- 3 A New Taxonomy
- 4 Responsibility and the Limitations of Interventionism

1 The Interventionist Theory

Causes and Interventions

"[C]ausal [...] claims are informed by our interest as practical agents in changing the world" (Woodward, 2003, 25).

"[T]he ordinary notions of cause and effect have a direct and essential connection with our ability to intervene in the world as agents" (Menzies and Price, 1993, 187).

Similar ideas can be found in older manipulationist theories (Gasking, 1955; von Wright, 1971). An exception is Collingwood.

"[...] the word cause expresses an idea relative to human action [...] *A cause is an event or state of things it is in our power to produce or prevent, and by producing or preventing which we can produce or prevent that whose cause it is said to be*" (Collingwood, 1938, 89, *emph. orig.*).

But note that:

"In the first [the proper] sense of the word cause, that which is caused is the free and deliberate act of a conscious and responsible agent" (Collingwood, 1938, 86).

Woodward's (2003) definition of total cause:

Let us say that X is a total cause of Y [...] if and only if there is some intervention on X alone (and no other variables) such that for some value of other variables besides X , this intervention on X will change the value of Y " (50).

This notion is particularly useful in experimentation, where scientists investigate the consequences of varying multiple parameters.

But sometimes we are interested in more specific causal claims.

Interventionists on Actual Causation

Many have tried to specify what actual causation is.

(Lewis, 1973; Pearl, 2009; Hitchcock, 2001; Woodward, 2003; Halpern and Pearl, 2005; Hitchcock, 2007; Halpern and Hitchcock, 2015)

Why are we interested in such claims? In particular, why can't we make do with causal models?

Hitchcock and Knobe give an answer in interventionist terms:

"[W]hile causal structure identifies all of the factors that could be manipulated (either singly or in combination) to effect change in the outcome, the actual causes are the factors that should be manipulated" (590).

Hitchcock (2017) specifies the point regarding combinations of interventions in contexts with complex causal structure where intervening agents follow multiple goals.

The pen vignette (Knobe and Fraser, 2008):

Professor Smith and an administrative assistant both take one of the last two pens on the receptionist desk. Later that day the receptionist cannot take an important message because there are no pens left. The administrative assistants are allowed to take pens, but faculty members are supposed to buy their own.

Test subjects identify Professor Smith's rather than the administrator's behavior as the cause of the problem.

Counterfactuals

1. Descriptive claim

Some counterfactuals are considered more relevant than others: "when something abnormal occurs, [people] regard as relevant counterfactuals those that involve something more normal having occurred instead" (597).

Norms

Norms can be statistical and prescriptive (moral, legal, or functional).

The different norms do not play distinct roles, instead "people classify each counterfactual as having a single overall degree of "normality"" (Hitchcock and Knobe, 2009, 598).

cf. also (Hart and Honoré, 1959; McGrath, 2005; Halpern and Hitchcock, 2015)

2. Evaluative claim:

it is useful to think about

- statistically normal counterfactuals because they describe situations **we will encounter more frequently**
- prescriptively normal counterfactuals because they describe situations **we try to bring about**

1. Descriptive claim

The relevant counterfactuals correspond to the factors we identify as actual causes and as suitable targets of intervention.

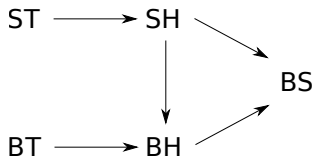
2. Evaluative claim

Actual causes are suitable targets of interventions.

- **statistical**: fixing the statistically abnormal is a generalizable strategy of intervention.
- **moral**: intervening on the morally bad (rather than the morally good) conditions in order to prevent a bad outcome increases the overall good.
- **functional**: intervening on dysfunctional elements (e.g. of a machine) in order to resolve an overall malfunctioning is more efficient than introducing other dysfunctional elements

Combining Interventions

Late pre-emption



$$SH = ST$$

$$BH = BT \wedge \neg SH$$

$$BS = SH \vee BH.$$

Intervening on the actual cause is not sufficient. We need an intervention that "clamps" BH, that is, fixes BH at its default value.

2 Criticism

Focussing on descriptive claims regarding pen-vignette like studies

- Norm-related folk intuitions of causation should not be included in a (descriptive) philosophical theory of causation (e.g. Mill (1843); Lewis (1973); Hall (2004); Rose (2017)).
- Pen-vignette like studies track the influence of norms on judgments of accountability rather than causation (Samland and Waldmann, 2016).
- Normative considerations concern only certain high-level and language-driven forms of causal reasoning (Danks et al., 2014).

Including criticism regarding functional and normative claims

- Distinguishing agent-level and population-level typicality supports the "responsibility view" (Sytsma et al., 2012).
- Criticism from the perspective of the Culpable Control Model (Alicke et al., 2011): (1) person as engineer vs person as evaluator, (2) causal judgements are portrayed as overly rational.

3 A New Taxonomy

Disambiguating Actual Causation

- actual causation as token causation
describes relations between token events at particular places and times
- actual causes vs potential causes
describes causes that *do*—rather than merely *can*—bring about effects
- actual causes vs background conditions
describes the most salient factors in bringing about an effect

The Taxonomy

AC1 actual causes vs potential causes

correspond to disjunctive causal structure: $Y = X_1 \vee \dots \vee X_n$.

AC2 actual causes vs background conditions

correspond to conjunctive causal structure: $Y = X_1 \wedge \dots \wedge X_n$.

Both kinds of claims can be made with regard to token and type events.

	AC1	AC2
token	Suzy's throwing	short-circuit
type	*	short-circuits

*e.g. cancer, redundancy in genetic regulatory systems (Mitchell, 2008)

Is this a new taxonomy? (Baumgartner and Fenton-Glynn, 2013; Hall, 2004).

The relation between AC1 and AC2

A case of AC2 can be turned into a case of AC1 (and vice versa) depending on the value of the effect variable that is to be explained.

Example: “why is there a fire?” vs. “why isn't there a fire?”

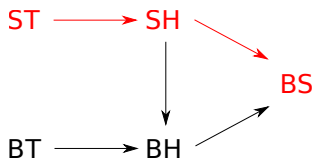
$$F = S \wedge O \text{ (AC2)}$$

$$\neg F = \neg(S \wedge O) \Leftrightarrow \neg F = \neg S \vee \neg O \text{ (AC1).}$$

Claims of AC1 often imply claims of AC2

1. In disjunctive structure ($Y = X_1 \vee \dots \vee X_n$) the X_i often are not sufficient. For example: strictly speaking Suzy's throwing is not sufficient for causing the bottle's shattering.
2. Late pre-emption cases involve conjunctive structure.

The relation between AC1 and AC2

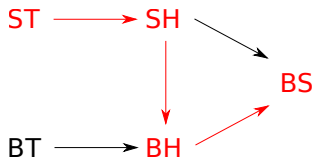


$ST = 1$ and $BT = 1$

$SH = ST$

$BH = BT \wedge \neg SH$

$BS = SH \vee BH.$



$ST = 0$ and $BT = 1$

$SH = ST$

$BH = BT \wedge \neg SH$

$BS = SH \vee BH.$

Going beyond the model

Actual causation describes kinds of reasoning that are not covered by interventionists causal models.

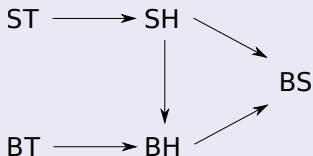
Note: Whether a claim of actual causation goes beyond the model depends both on the claim and on the model.

AC2: Actual Causes vs Background

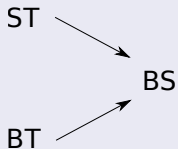
We need to distinguish causes from background conditions only if the causal model contains background conditions. There are good reasons not to include them (Halpern and Hitchcock, 2010; Blanchard and Schaffer, 2017).

AC1: Actual Causes vs Potential Causes

In the detailed Billy-and-Suzy model with assigned values the claim of actual causation does *not* go beyond the model. But the claim of actual causation does go beyond the simple model (which the interventionist should prefer, as we will see later).



$$\begin{aligned}SH &= ST \\BH &= BT \wedge \neg SH \\BS &= SH \vee BH.\end{aligned}$$



$$BS = SH \vee BH.$$

4 Responsibility and the Limitations of Interventionism

Overview

4.1 Token Causal Claims

4.2 Late Pre-emption (Type causal claims of AC1)

4.3 Early Pre-emption (Type causal claims of AC1)

Two points to clarify

1. Problems for interventionists
2. Advantages of incorporating responsibility

Two minimal restrictions on responsibility (in place of a full theory...)

- We need moral responsibility instead of merely causal responsibility.
- Moral responsibility as moral influence (as in e.g. Schlick (1930); Vargas (2008)) will not help.

Token causal claims

cause	effect	
past	past	<i>a caused b</i>
past	future	<i>a causes b</i>
future	future	<i>a will cause b</i>

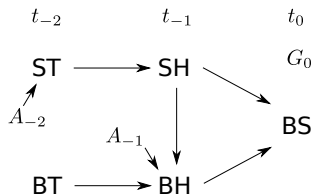
Many token claims of actual causation (AC1 and AC2) are past-past claims. Unless they can be generalized they do not indicate suitable targets for future intervention.

Yet, they can be important for liability claims.

Examples can be found in law and history.

Type causal claims of AC1 - Late Pre-emption

Hitchcock (2017) argues that actual causation identifies path-specific effects. Knowledge of these effects can be useful in complex cases where we reason backward from the goal.

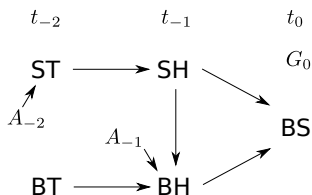


$$\begin{aligned}SH &= ST \\BH &= BT \wedge \neg SH \\BS &= SH \vee BH.\end{aligned}$$

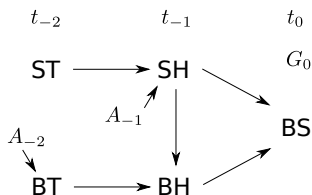
"[S]uppose I want the window to remain intact. If my only option is to intervene on Suzy, my goal will be thwarted. [...] However, if I can intervene on both Billy and Suzy, I can succeed. In order to arrive at the right combination of interventions, I need to know how Suzy's action will work in combination with Billy's action" (124).

Type causal claims of AC1 - Late Pre-emption

Strategy 1



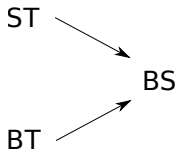
Strategy 2



Is there any reason to prefer strategy 1 over strategy 2?

Type causal claims of AC1 - Late Pre-emption

Late pre-emption and overdetermination



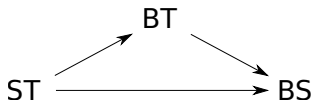
$$BS = SH \vee BH.$$

The interventionist should prefer the simple model as a description of the Billy-Suzy case because it is just as informative, is simpler and does not involve non-causal relations between variables (Halpern and Hitchcock, 2010; Woodward, 2016).

By contrast, distinguishing overdetermination from late pre-emption is relevant with regard to responsibility (moral luck).

Type causal claims of AC1 - Early Pre-emption

Early pre-emption



$$BT = \neg ST$$

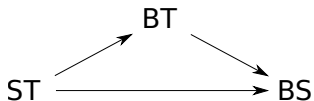
$$BS = ST \vee BT.$$

The claim of actual causation leaves vastly underdetermined what is necessary for preventing the effect.

By contrast, ascription of responsibility is not affected by pre-empted factors.

Type causal claims of AC1 - Early Pre-emption

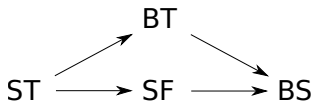
Early pre-emption



$$BT = \neg ST$$

$$BS = ST \vee BT.$$

Include intermediate variable.

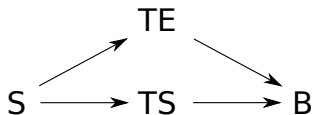
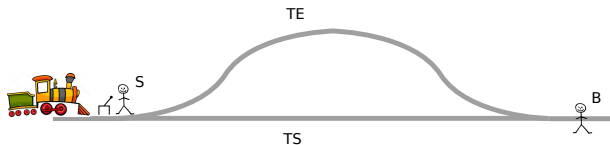


$$SF = ST$$

$$BT = \neg ST$$

$$BS = BT \vee SF.$$

Type causal claims of AC1 - Early Pre-emption



$$TS = S$$

$$TE = \neg S$$

$$B = TE \vee TS.$$

Why (from the interventionist perspective) is Suzy a cause of the bottle's shattering but not of Billy's death?

Assume that responsibility is transitive wrt variables that describe objects but not wrt to variables that describe persons. Then including responsibility helps explaining the difference.

When analysing the function of actual causation and corresponding causal reasoning, we need to keep in mind that actual causation can mean very different things.

It is difficult to maintain a purely interventionist view as regards

- token claims of actual causation that concern past-past relations,
- (at least some) late pre-emption cases concerning type relations,
- (at least some) early pre-emption cases concerning type relations.

Incorporating responsibility into functional accounts helps in these cases.

- Alicke, Mark D., Rose, David, and Bloom, Dori. Causation, norm violation, and culpable control. *The Journal of Philosophy*, 108(12): 670–696, 2011.
- Baumgartner, Michael and Fenton-Glynn, Luke. Introduction to special issue on 'actual causation'. *Erkenntnis*, 78(S1):1–8, 2013.
- Blanchard, Thomas and Schaffer, Jonathan. Cause without default. In Helen Beebe, Huw Price, Christopher Hitchcock, editor, *Making a Difference*, pages 175–214. Oxford University Press, 2017.
- Collingwood, Robin George. On the so-called idea of causation. *Proceedings of the Aristotelian Society*, New Series 38:85–112, 1938.
- Danks, David, Rose, David, and Machery, Edouard. Demoralizing causation. *Philosophical Studies*, 171:251–277, 2014.
- Gasking, Douglas. Causation and recipes. *Mind*, 64(256):479–487, 1955.
- Hall, Ned. Two concepts of causation. In Collins, John, Hall, Ned, and Paul, Laurie, editors, *Causation and Counterfactuals*, pages 225–276. MIT Press, 2004.
- Halpern, Joseph and Hitchcock, Christopher. Actual causation and the art of modeling. In Dechter, Geffner and Halpern, editors, *Heuristics, Probability, and Causality: A Tribute to Judea Pearl*, pages 383–406. College Publications, 2010.
- Halpern, Joseph Y. and Hitchcock, Christopher. Graded causation and defaults. *The British Journal of the Philosophy of Science*, 66: 413–457, 2015.
- Halpern, Joseph Y. and Pearl, Judea. Causes and explanations: A structural-model approach. part i: Causes. *British Journal for the Philosophy of Science*, 56(4):843–887, 2005.
- Hart, H. L. A. and Honoré, A. M. *Causation in the Law*. Clarendon Press, Oxford, 1959.
- Hitchcock, Christopher. The intransitivity of causation revealed in equations and graphs. *The Journal of Philosophy*, 98(6):273–299, 2001.
- Hitchcock, Christopher. Prevention, preemption, and the principle of sufficient reason. *Philosophical Review*, 66:495–532, 2007.
- Hitchcock, Christopher. Actual causation: What's the use? In *Making a Difference*, pages 116–131. OUP, 2017.
- Hitchcock, Christopher and Knobe, Joshua. Cause and norm. *The Journal of Philosophy*, 106(11):587–612, 2009.
- Knobe, Joshua and Fraser, Ben. Causal judgment and moral judgment: Two experiments. In Sinnott-Armstrong, Walter, editor, *Moral Psychology, vol. 2: The Cognitive Science of Morality: Intuition and Diversity*, pages 441–448. MIT Press, 2008.
- Lewis, David. Causation. *Journal of Philosophy*, 70(17):556–567, 1973.
- McGrath, Sarah. Causation by omission: A dilemma. *Philosophical Studies*, 123(1):125–148, 2005.
- Menzies, Peter and Price, Huw. Causation as a secondary quality. *The British Journal for the Philosophy of Science*, 44:187–203, 1993.
- Mill, John Stuart. *A System of Logic*. John W. Parker, 1843.
- Mitchell, Sandra D. Exporting causal knowledge in evolutionary and developmental biology. *Philosophy of Science*, 75:697–706, 2008.
- Pearl, Judea. *Causality. Models, Reasoning, and Inference*. Cambridge University Press, 2 edition, 2009.
- Rose, David. Folk intuitions of actual causation: a two-pronged debunking explanation. *Philosophical Studies*, 174:1323–1361, 2017.
- Samland, Jana and Waldmann, Michael. How prescriptive norms influence causal inferences. *Cognition*, 156:164–176, 2016.
- Schlick, Moritz. *Fragen der Ethik*. Springer, Vienna, 1930.
- Sytsma, Justin, Livengood, Jonathan, and Rose, David. Two types of typicality: Rethinking the role of statistical typicality in ordinary causal attributions. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 43:814–820, 2012.
- Vargas, Manuel. Moral influence, moral responsibility. In Trakakis, Nick and Cohen, Daniel, editors, *Free Will and Moral Responsibility*, pages 90–122. Cambridge Scholars Press, 2008.
- von Wright, Georg Henrik. *Explanation and Understanding*. Cornell University Press, 1971.
- Woodward, James. *Making Things Happen*. Oxford University Press, Oxford, 2003.
- Woodward, James. The problem of variable choice. *Synthese*, 193:1047–1072, 2016.