

Lewis Causation is a Special Case of Spohn Causation

Franz Huber

Lewis ([1973a]) famously defined event c to be causally relevant to event e in world w just in case $O(c)$ and $O(e)$ as well as $\neg O(c) \Box \rightarrow \neg O(e)$ are true in w , where $O(c)$ and $O(e)$ say that c occurs and that e occurs, respectively. For him, this implied that $O(c) \Box \rightarrow O(e)$ is true, because he was assuming the underlying logic to be **VC**, which validates $(\alpha \wedge \gamma) \supset (\alpha \Box \rightarrow \gamma)$. If one works with the weaker system **V**, the causal relevance of event c to event e has to be defined as follows:

- (1) both c and e occur in w ; i.e. $O(c)$ and $O(e)$ are true in w ;
- (2) if c had not occurred, e would not have occurred either in w ; i.e. $\neg O(c) \Box \rightarrow \neg O(e)$ is true in w ;
- (3) if c had occurred, e would have occurred as well in w ; i.e. $O(c) \Box \rightarrow O(e)$ is true in w ;

A typical semantics for counterfactual conditionals says that ‘if A were the case, C would be the case’, $\alpha \Box \rightarrow \gamma$, is true if and only if C is true in all the closest A -worlds (Stalnaker [1968]; Lewis [1973b]). Closeness or distance is spelt out in terms of a relation between possible worlds. In another paper (Huber [unpublished]), I have defended a rank-theoretic semantics for counterfactual conditionals based on a principle called the *Royal Rule*, linking objective modalities and subjective beliefs. This principle implies that distance has the structure of a ranking function.

Ranking functions have been developed in (Spohn [1988]). They can be defined as functions r from an algebra of propositions \mathcal{A}_W over a set of possible worlds W into the set of extended natural numbers $\mathbb{N} \cup \{\infty\}$ such that the tautological proposition W is assigned rank 0, $r(W)=0$, the contradictory proposition \emptyset is assigned rank ∞ , $r(\emptyset)=\infty$, and the rank of a disjunction $A \cup B$ equals the minimum of the ranks of the disjuncts A and B , $r(A \cup B)=\min\{r(A), r(B)\}$.

According to this rank-theoretic semantics a counterfactual conditional $\alpha \Box \rightarrow \gamma$ is true at world w in model $(W, \mathcal{A}_W, (r_w)_{w \in W}, \varphi)$ if and only if $\varphi(\alpha)^w \subseteq \varphi(\gamma)$. Here, W is the set of all possible worlds, \mathcal{A}_w is an algebra of

propositions over W , $(r_w)_{w \in W}$ is a family of ranking functions defined on \mathcal{A}_w , and φ is an interpretation function on the underlying language \mathcal{L} . $\varphi(\alpha)$ is the set of worlds w from W in which the sentence α from the language \mathcal{L} is true, and $\varphi(\alpha)^{r_w}$ is the set of r_w -minimal elements of $\varphi(\alpha)$ —that is, the set of worlds in which α is true and that are closest to w from the point of view of r_w .

The system **V** is sound and complete with respect to this rank-theoretic semantics. If the model $(W, \mathcal{A}_W, (r_w)_{w \in W}, \varphi)$ is weakly [respectively strongly] centered in the sense that $r_w(\{w'\}) = 0$, if [respectively if and only if] $w = w'^1$, then the system **VW** [respectively **VC**] is sound and complete with respect to the resulting weakly [respectively strongly] centered rank-theoretic semantics. **V**, **VW**, and **VC** are axiomatized in (Lewis [1973b], Chapter 6).

In terms of the rank-theoretic semantics for counterfactuals the last two clauses of Lewis' definition of causal relevance translate into

$$\begin{aligned} 2'. \quad & r_w(O(e)|\overline{O(c)}) > r_w(\overline{O(e)}|\overline{O(c)}), \\ 3'. \quad & r_w(\overline{O(e)}|O(c)) > r_w(O(e)|O(c)), \end{aligned}$$

provided $r_w(\overline{O(c)})$ and $r_w(O(c))$ are finite. Causation itself is then defined as the transitive closure of causal relevance.

There are several well-known counterexamples to Lewis' counterfactual theory of causation, and Lewis has since refined his account (Lewis [1986], [2000]). The interested reader is referred to the collection of articles in (Collins *et al.* [2004]). In a parallel effort, people have tried to provide a probabilistic theory of causation, starting with Suppes ([1970]) up to the sophisticated work of Spirtes *et al.* ([2000]). Pearl ([2009]) and Woodward ([2003]) pay homage to both traditions, although the former leans more towards the probabilistic, and the latter more towards the counterfactual paradigm.

Spohn ([1983], [2006]) develops a theory of causation in the tradition of the probabilistic paradigm except that he is working with ranking functions instead of probability measures. He is able to elegantly deal with the problems besetting Lewis' account—most notably causation by overdetermination and causation by preemption—but the one big problem his account faces is that it is *subjective*.

More precisely, causal relevance—actually (Spohn [2006]) starts with the narrower notion of direct causation—and causation are only defined relative to an epistemically interpreted ranking function. Spohn is aware of this problem and has developed a way to partly 'objectify' ranking functions by associating them with propositions (Spohn [1993]). However, his attempt at objectification is more complicated than successful. In particular, the crucial notions of direct causation and causation are only conditionally objectifiable.

¹ If the algebra \mathcal{A}_w is not rich enough to contain all singleton propositions $\{w'\}$, the formulation is more complicated.

Given a semantics of counterfactuals in terms of objective ranking functions, one is able to interpret Spohn's account of causation in terms of these objective ranking functions. It is then interesting to consider the relation between this objective version of Spohn's account of causation and the version of Lewis' account of causation one obtains from the rank-theoretic semantics for counterfactuals.

Ignoring some details—in particular, Spohn ([2006]) assumes that effects never predate their causes—we can paraphrase Spohn's definition of direct causation of event e by event c in world w as follows: both c and e occur in w ; i.e. $O(c)$ and $O(e)$ are true in w , and Spohn's Inequality holds:

$$\begin{aligned} Ra(\overline{O(e)}|O(c) \cap H_w) - Ra(O(e)|O(c) \cap H_w) \\ > Ra(\overline{O(e)}|\overline{O(c)} \cap H_w) - Ra(O(e)|\overline{O(c)} \cap H_w) \end{aligned}$$

Here H_w is the complete history of world w up to right before the effect e , but excluding the cause c (Spohn's framework allows him to give a precise formulation of this clause). As before, causation itself is then defined as the transitive closure of direct causation.

For Spohn, Ra is an ideally rational agent's prior ranking function. According to the rank-theoretic semantics counterfactuals receive truth values relative to presuppositions or contexts (Stalnaker [1974], [1998]). If we assume, as is plausible, that the relevant context for direct causation is the complete history up to right before the effect but excluding the cause, H_w , and if we further assume that the agent can know that c causes e only if she is certain of the relevant modalities T_w , i.e. $Ra(\cdot) = Ra(\cdot|T_w)$, it follows from the Royal Rule that $r_w(\cdot) = Ra(\cdot|H_w \cap T_w)$.

Given these assumptions, Lewis causation turns out to be a special case of Spohn causation, because causal relevance is a special case of direct causation. Suppose c is causally relevant to e in w in the sense of Lewis. Then the left-hand side of Spohn's Inequality is positive and the right hand-side of it is negative. Thus, causal relevance implies direct causation. The converse is not true, because Spohn's Inequality may hold if both sides are, say, positive, but the right-hand side is greater than the left-hand side. According to Spohn ([2006], Section 5) this happens in cases of causal overdetermination and trumping.

Acknowledgements

I am grateful to Wolfgang Spohn for helpful comments on an earlier version of this article. My research was supported by the German Research Foundation through its Emmy Noether program.

Formal Epistemology Research Group
Zukunftskolleg and Department of Philosophy
PO Box 216
University of Konstanz, Germany
franz.huber@uni-konstanz.de

References

- Collins, J., Hall, N. and Paul, L. A. (eds) [2004]: *Causation and Counterfactuals*, Cambridge, MA: MIT Press.
- Huber, F. [unpublished]: 'New Foundations for Counterfactuals'.
- Lewis, D. K. [1973a]: 'Causation', *Journal of Philosophy*, **70**, pp. 556–67.
- Lewis, D. K. [1973b]: *Counterfactuals*, Cambridge, MA: Harvard University Press.
- Lewis, D. K. [1986]: 'Postscripts to "Causation"', in his *Philosophical Papers II*, Oxford: Oxford University Press, pp. 172–213.
- Lewis, D. K. [2000]: 'Causation as Influence', *Journal of Philosophy*, **97**, pp. 182–97.
- Pearl, J. [2009]: *Causality: Models, Reasoning, and Inference*, 2nd edition, Cambridge: Cambridge University Press.
- Spirtes, P., Glymour, C. and Scheines, R. [2000]: *Causation, Prediction, and Search*, 2nd edition, Cambridge, MA: MIT Press.
- Spohn, W. [1983]: *Eine Theorie der Kausalität*, Unpublished Habilitation thesis, Munich: LMU Munich.
- Spohn, W. [1988]: 'Ordinal Conditional Functions: A Dynamic Theory of Epistemic States', in W. L. Harper and B. Skyrms (eds), *Causation in Decision, Belief Change, and Statistics II*, Dordrecht: Kluwer, pp. 105–34.
- Spohn, W. [1993]: 'Causal Laws are Objectifications of Inductive Schemes', in J. Dubucs (ed.), *Philosophy of Probability*, Dordrecht: Kluwer, pp. 223–52.
- Spohn, W. [2006]: 'Causation: An Alternative', *British Journal for the Philosophy of Science*, **57**, pp. 93–119.
- Suppes, P. [1970]: *A Probabilistic Theory of Causality*, Amsterdam: North-Holland Publishing Company.
- Stalnaker, R. C. [1968]: 'A Theory of Conditionals', in N. Rescher (ed.), *Studies in Logical Theory*, American Philosophical Quarterly Monograph Series 2, Oxford: Blackwell, pp. 98–112.
- Stalnaker, R. C. [1974]: 'Pragmatic Presuppositions', in M. K. Munitz and P. Unger (eds), *Semantics and Philosophy*, New York: New York University Press, pp. 197–213.
- Stalnaker, R. C. [1998]: 'On the Representation of Context', *Journal of Logic, Language, and Information*, **7**, pp. 3–19. Reprinted in R. Stalnaker [1999]: *Context and Content*, Oxford: Clarendon Press, pp. 96–113.
- Woodward, J. [2003]: *Making Things Happen: A Theory of Causal Explanation*, Oxford: Oxford University Press.