

Introduction to Special Issue on 'Actual Causation'

Michael Baumgartner[†] and Luke Glynn[‡]Published online by *Erkenntnis* in March 2013

DOI: 10.1007/s10670-013-9441-8

URL: <http://link.springer.com/article/10.1007/s10670-013-9441-8>**Please cite published version only.**

An *actual cause* of some token effect is itself a (distinct) token event (or fact, or state of affairs, ...) that helped to bring about that effect. The notion of an actual cause is different from that of a *potential cause* – for example a pre-empted backup – which had the capacity to bring about the effect, but which wasn't in fact operative on the occasion in question. Sometimes *actual causes* are also distinguished from mere *background conditions*: as when we judge that the struck match was a cause of the fire, while the presence of oxygen was merely part of the relevant background against which the struck match operated. Actual causation is also to be distinguished from *type causation*: actual causation holds between token events in a particular, concrete scenario; type causation, by contrast, holds between event kinds in scenario kinds.¹

Modern philosophical study of actual causation dates back to Hume (1739-1740), with Mill (1843), Mackie (1965), and Lewis (1973) making field-defining contributions to the subject. While the late twentieth century saw vibrant debate concerning actual causation, there has been a particular surge of interest since the turn of the millennium. One of the key reasons for this has been an increased awareness among philosophers of the availability of formal tools – in particular *structural equation models* and *causal graphs* – that can fruitfully be brought to bear in the development of novel solutions to traditional problems confronting attempts to analyze actual causation. Books by Spirtes, Glymour, and Scheines (1993/2000) and Pearl (2000/2009) have been key in bringing these formal tools to the attention of philosophers, with seminal attempts to analyze actual causation using these tools having been made by Pearl (2000, Ch. 10), Halpern and Pearl (2001), Hitchcock (2001), and Woodward (2003, esp. pp. 74-86). Over the subsequent decade or so, this tradition has been greatly enriched by further contributions, including those of Menzies (2004), Halpern and Pearl (2005), Glymour and Wimberly (2007), Hitchcock (2007a), Halpern (2008), and Halpern and Hitchcock (2010, ms.).

Though these structural equation approaches to the analysis of actual causation have been highly influential, there continues to be a healthy diversity of alternative approaches that are

[†] University of Osnabrück, Institute of Cognitive Science, Albrechtstrasse 28, 49076 Osnabrück, Germany. Email: <michael.baumgartner@uni-osnabrueck.de>.

[‡] California Institute of Technology, HSS Division, MC 101-40, 1200 E. California Blvd., Pasadena, CA 91125, U.S.A. Email: <lglynn@caltech.edu>.

¹ The term 'actual causation' became firmly established in the philosophical lexicon with Pearl (2000). It has, however, been used extensively in the tort law literature for some time (see e.g. Wright 1985). The term refers to a relation that philosophers often used to (and sometimes still do) call 'token causation'. In his contribution to this volume, Christopher Hitchcock argues that 'actual causation' is less misleading than 'token causation' when deployed as a term for that token-level causal relation that has been the standard target of philosophical analysis. This is because, Hitchcock argues, the standard target of philosophical analysis is not the only kind of token-level causal relation. In their contributions, Max Kistler and David Danks also suggest that actual and token causation should not be equated.

pursued. Structural equations are standardly understood as representing relations of non-backtracking counterfactual dependence between events (or rather, variables representing those events).² Such approaches can thus be seen as one branch of the tradition, initiated by Lewis (1973), of attempts to analyze actual causation in counterfactual terms. But, on the one hand, not all of those who have implemented the structural equation framework to analyze causation have done so against the background of a counterfactual understanding of causation,³ and, on the other, not all of those who have recently pursued counterfactual accounts of actual causation have done so within the structural equation framework (see, e.g., Yablo 2004, and Ramachandran 2004).⁴ There is a healthy diversity of rivals to the counterfactual approach. These include *process* (or *transference*) accounts (such as that developed by Dowe 2000, 2004), *mechanisms* approaches (which include the seminal contribution of Machamer, Darden, and Craver 2000), probabilistic approaches (including Kwart 2004 and Glynn 2011), regularity approaches (including Graßhoff and May 2001, Baldwin and Neufeld 2004, Strevens 2007, and Baumgartner 2008; cp. also Halpern 2008, Section 5), and realist approaches (including Mumford and Anjum 2011, or Chakravartty 2005).

Structural equations are a particular representational device: as noted, they are normally construed as representing relations of non-backtracking counterfactual dependence, relations to which actual causation is reducible according to the counterfactual analyst. Since they are readily able to represent rather subtle patterns of counterfactual dependence (including *de facto* or *explicitly non-foretracking* counterfactual dependence⁵), they are much vaunted for their heuristic value in facilitating the development of new and sophisticated counterfactual analyses of causation. Nevertheless, there is a lively discussion concerning the question of which is the most felicitous and fecund device for representing causal structures. Lewis (1986) first proposed 'neuron' diagrams as a device for representing the relations of counterfactual dependence in terms of which he sought to analyze causation. Hall (2007) discusses the merits, and some of the limitations, of their use, while Hitchcock (2007b, 2009) has argued for the superiority of structural equations and causal graphs. Alternative devices for representing causal relationships have been proposed by philosophers working in other traditions: for instance, Mumford and Anjum (2011) criticize neuron diagram representations as being ideologically biased: specifically, they argue that neuron diagrams are conducive to a Humean ontology, which they reject in favor of realism about causal powers. They argue that causal powers are better represented as *vectors*.

Another interesting aspect of the recent literature on actual causation concerns the question of whether the traditional philosopher's project of attempting to understand actual causation by providing putative necessary and sufficient conditions, which are then compared to intuitions concerning a battery of test cases, is a fruitful pursuit. Glymour *et al.* (2010) argue that such an enterprise is bound to fail. In particular, they argue that standard test cases are inevitably very simple, and constitute only a tiny proportion of a huge number of possible causal structures that cannot be surveyed by intuitions. What is worse, many interestingly novel causal features only appear in structures of considerably higher complexity than the standard test cases. (Neuron

² On the notion of non-backtracking counterfactual dependence – which is the variety of counterfactual dependence in terms of which philosophers have attempted to analyze actual causation – see Lewis (1979).

³ Indeed it is possible to construe structural equations as representing things other than counterfactual dependence relations: for example, Pearl (2000) speaks of them as representing *mechanisms*; while Baldwin and Neufeld (2004) speak of them as representing *law-like regularities*. They are thus found to be useful tools even by some who expound non-counterfactual accounts of actual causation (e.g. Hiddleston 2005, or Handfield *et al.* 2008).

⁴ Hall (2007) provides a counterfactual account of causation that he thinks can, but need not be, presented within the structural equation framework.

⁵ These terms are due respectively to Yablo (2004) and Hitchcock (2001).

diagram and, especially, structural equation representations are apt to suggest such novel features.) In place of the standard philosophical method – traceable to Plato – of conceptual (or metaphysical) *analysis*, they advocate a more mathematical, Euclidean, axiomatic method, consisting in the provision of principles that serve as reliable indicators for discovering actual causal relations. These principles or axioms are justified to the extent that they are fruitful in allowing for causal generalization, inference, control, etc. The idea is that such a set of axioms will reproduce the basic tenets of our causal reasoning, at the same time showing why it is useful for us to have a concept of actual cause that accords with these principles. Along the way, it is hoped that the content of our concept of actual causation will be illuminated. But the criterion of success in this project is not the provision of a reductive definition of actual cause that captures ordinary intuitions about all cases. The prescription of minor conceptual revisions, and explication rather than reduction, are not considered deadly sins on such an approach.⁶

As shall be explained below, each of these various strands of the current debate concerning actual causation is discussed in articles that this special issue comprises. The special issue is the result of a workshop on actual causation organized at the University of Konstanz in September 2010. The workshop was organized in the context of the 'Causality and Probability' project, which ran from 2009-2012. This was a collaborative project involving philosophers at Konstanz and at the *Institut d'Histoire et de Philosophie des Sciences et des Techniques* (IHPST) in Paris. The IHPST side of the project was directed by Jacques Dubucs and funded by *L'Agence Nationale de la Recherche* (ANR). The Konstanz side was directed by Wolfgang Spohn and funded by the *Deutsche Forschungsgemeinschaft* (DFG) under the grant SP279/15-1, which also funded the workshop. The editors of this special issue were both employed as post-doctoral researchers at Konstanz, within the context of the project, at the time of the workshop.

The workshop brought together a number of world-leading researchers who have made significant contributions to recent work on actual causation. The articles collected in this issue grew out of talks given at that workshop, and advance the debate on actual causation in a number of directions. What follows is a summary of the articles, with some connections drawn between them.

In the article entitled 'Metaphysically Reductive Causation', Ned Hall and L. A. Paul argue that the project of seeking a metaphysical reduction of actual causation is of significant philosophical value and interest. They defend the value and interest of such a project *vis-à-vis* some other projects that might be pursued in the philosophy of actual causation, such as non-reductive *explication* of the actual causal relation, Fregean-style *conceptual analysis*, or an analysis of the *function* that the concept of cause serves in the human cognitive economy. (The latter sort of project is pursued by David Danks in his contribution to this special issue.) Hall and Paul provide a set of rules, or methodological maxims, that they argue ought to guide the development of a metaphysically reductive account of causation. This allows them to provide a principled and systematic evaluation of the most significant accounts currently on the market – namely, regularity accounts, counterfactual accounts (including their structural equation variants), probabilistic accounts, and transference accounts – in terms of the extent to which they satisfy these principles. One of their maxims is that one ought not to smuggle causal facts into the analysis of one's analysis. In discussing counterfactual accounts, and their structural equation variants, Hall and Paul emphasize that this means, among other things, that the semantics for the relevant counterfactuals must not appeal to causal facts. Whether structural equation accounts can get away with a non-causal semantics for counterfactuals is an issue that is further explored in Luke Glynn's contribution to this special issue. In giving examples, Hall and Paul make extensive use of neuron diagrams, and explain the basic conventions governing such representations, as well as some of the potential pitfalls of using neuron diagrams to represent

⁶ On the virtues of such an approach, see also Spirtes *et al.* (2000, pp. 3-4) and Woodward (2003, Ch. 1).

causal structures.

In his contribution entitled 'Of Miracles and Interventions', Luke Glynn provides a detailed outline of the use of structural equations in representing counterfactual dependence relations, and also of the various accounts of actual causation that have been developed within the structural equation framework. Glynn examines an important foundational question concerning the appropriate semantics for the counterfactuals encoded by structural equations. A highly influential approach, due to Lewis (1979), has it that the counterfactuals relevant to causation are to be evaluated with respect to possible worlds in which their antecedents are realized by small *miracles*. This approach has been critiqued by Woodward (2003), whose rival, *interventionist*, semantics has gained a good deal of currency, especially among advocates of structural equation accounts of actual causation. Glynn defends Lewis's semantics against Woodward's criticisms, arguing that structural equation analyses of actual causation perform at least as well – in terms of extensional adequacy – when founded upon (a generalized version of) Lewis's semantics as when founded upon Woodward's. Moreover, Glynn argues that adopting the Lewisian semantics avoids certain circularities associated with Woodward's approach, and thus holds out the promise of reduction.

In the paper entitled 'The Interventionist Account of Causation and Non-Causal Association Laws', Max Kistler also critiques Woodward's interventionist version of a structural equation analysis. This time the focus of the critique is on the extensional adequacy (rather than the non-reductivity) of Woodward's approach. Specifically, it is argued that Woodward's account does not have the resources for distinguishing between genuine causal relations and non-causal nomological associations. As an example of a non-causal association law, Kistler considers the 'magnetic stirrer law', which relates the *angular momentum* (\mathbf{L}) of a magnetic stirrer to its *magnetic moment* ($\boldsymbol{\mu}$). The variables \mathbf{L} and $\boldsymbol{\mu}$ are *generic*, in the sense that many different systems at many different times may each possess values for these variables. But Kistler observes that we can construct specific variables $\mathbf{L}(s, t)$ and $\boldsymbol{\mu}(s, t)$ that represent the values of \mathbf{L} and $\boldsymbol{\mu}$ for a particular (token) magnetic stirrer, s , at a particular time, t . Kistler argues that an experimental manipulation of the value of the particular stirrer's angular momentum, $\mathbf{L}(s, t)$ (that is, an intervention upon this variable), would (by the association law) change the value of its magnetic moment, $\boldsymbol{\mu}(s, t)$, and that an intervention on its magnetic moment, $\boldsymbol{\mu}(s, t)$, would change the value of its angular momentum, $\mathbf{L}(s, t)$. It follows, on Woodward's interventionist account, that $\mathbf{L}(s, t)$ is a *cause* of $\boldsymbol{\mu}(s, t)$ and *vice versa*. But, Kistler observes, this is incompatible with the asymmetry of (token-level) causation. Indeed, Kistler points out, if we additionally assume that causation is transitive, we will get the absurd result that $\boldsymbol{\mu}(s, t)$ and $\mathbf{L}(s, t)$ each cause *themselves*. Consequently, Kistler argues that Woodward's interventionist analysis does not provide a sufficient condition for causation, but merely provides a condition that is (necessary and) sufficient for nomological dependence, a relation which is not asymmetric and which includes (not just causal dependence but also) dependence via a non-causal association law. Kistler ends his paper by suggesting that token-level causation between two variables might be analyzable in terms of the notion of nomological dependence *together* with that of a causal process.

In his article 'A Regularity Theoretic Approach to Actual Causation', Michael Baumgartner develops a sophisticated regularity account of both type causation and actual causation, arguing that his account overcomes problems associated with classic regularity accounts, such as those of Mackie (1965) and Wright (1985). Baumgartner also argues that his account performs at least as well – in extensional adequacy terms – as sophisticated counterfactual accounts of actual causation that draw upon structural equations models. Indeed, he argues that his account is superior to certain structural equation analyses in its treatment of cases of so-called *switching* and *short-circuiting*. He also defends his account against a putative counterexample to regularity analyses offered by Hall (2004, p. 263). In fact, Baumgartner is able to appeal to many of the innovations and insights of the structural equations literature – such as

Hitchcock's (2001) notion of an *active route* and the deployment by Hitchcock (2007), Halpern (2008), and others of *normality* or *typicality* considerations – while setting them in the context of a regularity approach. Baumgartner argues that, in addition to performing at least as well as the best structural equation accounts, his regularity theoretic approach has the advantage of not appealing – as standard structural equation approaches do – to such baroque metaphysical and conceptual resources as *non-actual possible worlds* and *explicitly non-foretracking counterfactuals*. Thus, he argues, his account is preferable on grounds of parsimony. Indeed, Baumgartner argues that his account is thoroughly empiricist and free of modal notions: it appeals only to material conditionals, Boolean minimization procedures, and a (non-modal) stability condition concerning the behavior of causal models under expansions of the sets of analyzed factors.

The starting point of the contribution entitled 'Functions and Cognitive Bases for the Concept of Actual Causation' by David Danks is an acknowledgement of the challenge that the arguments of Glymour *et al.* (2010) – which were described above – pose for traditional approaches to analyzing actual causation. In response, Danks advocates a different approach: namely one that seeks to examine the principles that generate judgments about actual causation. The aim of such an approach is to shed light upon the concept of actual causation, without (necessarily) providing necessary and sufficient conditions for its application. Danks advocates two (complementary) methods – each distinct from the standard practice of intuition-mining, and surveying test cases – that may fruitfully be pursued with a view to eliciting such general principles. The first is an analysis of the functional role that the concept of causation plays in the human cognitive economy: for example, the role that it plays in supporting explanation, prediction, and control. The idea is to attempt to cast light upon the content of our concept of actual causation by examining the role the concept plays for us and why we have the concept in the first place. The second method that Danks advocates involves an exploration of the cognitive mechanisms that underlie our experiences of, and judgments and reasoning about, actual causation. The idea is that we can thereby learn something about the scope and application of our concept of actual causation. Danks takes some initial steps toward implementing these two methods (the second of which he pursues by means of a detailed analysis of results from experimental psychology concerning actual causal perception). He argues that the results, even of these preliminary attempts to implement his two proposed methods, are some interesting insights into the concept of actual causation: specifically, Danks argues, they show that it is a concept of a relation that is graded, context-sensitive, and extrinsic.

Finally, in his article entitled 'What is the "Cause" in Causal Decision Theory?', Christopher Hitchcock is concerned with the role that causation plays in other philosophically interesting phenomena – rational decision-making, in particular. Noting that counterfactual theories of causation face serious problems with cases of preemption and (symmetric) overdetermination, Hitchcock points out that we should expect that such cases will pose problems for counterfactual theories of various other causally-involved phenomena too. Hitchcock illustrates this by describing how simple counterfactual theories of three such phenomena – memory, free will/moral responsibility, and dispositions – have succumbed to counterexamples involving pre-emption and overdetermination. (Examples could obviously be multiplied: see for instance Lewis's (1980, pp. 248-9) rather odd attempt to dismiss what seems to be a clear preemption-involving counterexample to his own counterfactual analysis of *seeing*.) Yet Hitchcock argues that rational decision is an exception to this. Causal decision theory takes rational decision-making to be a causal notion and, on a standard formulation, analyzes rational decision-making in terms of counterfactuals. This would lead us to expect counterexamples to causal decision theory involving cases of pre-emption and symmetric overdetermination. But, Hitchcock argues, we do not find such counterexamples. Hitchcock concludes that the 'cause' in causal decision theory is a causal notion that admits of a simple counterfactual analysis, and therefore can't be the same causal notion that is the standard target for philosophical analyses, including standard counterfactual analyses. That is, it can't be the notion of *actual cause*. Since the

notion of cause involved in causal decision theory is nevertheless (like the notion of actual causation) a *singular*, or *token*-level causal notion, Hitchcock holds that it would be wrong to simply equate *actual* causation with *token* causation. Rather, what is distinctive about actual causation is that it is *retrospective*: it is the notion that is involved in after-the-fact attributions of causal responsibility. The notion of causation that is involved in causal decision theory, by contrast, is *prospective*: it involves reasoning from causes to their effects.⁷ Hitchcock explains why both of these causal concepts are valuable to us.

References

- Baldwin, R. A. & Neufeld, E. (2004). The structural model interpretation of the NESS Test. *Advances in artificial intelligence, Lecture notes in computer science*, 3060, 292-307
- Baumgartner, M. (2008). Regularity theories reassessed. *Philosophia*, 36, 327-354
- Chakravartty, A. (2005). Causal realism: Events and processes. *Erkenntnis*, 63, 7-31
- Dowe, P. (2000). *Physical causation*. (Cambridge: Cambridge University Press)
- Dowe, P. (2004). Chance-lowering causes. (In P. Dowe & P. Noordhof (Eds.), *Cause and chance: Causation in an indeterministic world* (pp. 28-38). London: Routledge.)
- Glymour, C., Danks, D., Glymour, B., Eberhardt, F., Ramsey, J., Scheines, R., Spirtes, P., Teng, C. M., & Zhang, J. (2010). Actual causation: A stone soup essay. *Synthese*, 175, 169-192
- Glymour, C. & Wimberly, F. (2007). Actual causes and thought experiments. (In J. K. Campbell, M. O'Rourke, & H. Silverstein (Eds.), *Causation and explanation* (pp. 43-67). Cambridge, MA: MIT Press.)
- Glynn, L. (2011). A probabilistic analysis of causation. *British Journal for the Philosophy of Science*, 62, 343-392
- Graßhoff, G. & May, M. (2001). Causal regularities. (In W. Spohn, M. Ledwig, and M. Esfeld (Eds.), *Current issues in causation* (pp. 85-114). Paderborn: Mentis.)
- Hall, N. (2004). Two concepts of causation. (In J. Collins, N. Hall & L. A. Paul (Eds.), *Causation and counterfactuals* (pp. 225-276). Cambridge, MA: MIT Press.)
- Hall, N. (2007). Structural equations and causation. *Philosophical Studies*, 132, 109-136
- Handfield, C., Twardy, R., Korb, K. B. & Oppy, G. (2008). The metaphysics of causal models: Where's the biff? *Erkenntnis*, 68, 149-68
- Halpern, J. Y. (2008). Defaults and normality in causal structures. (In G. Brewka & J. Lang (Eds.), *Proceedings of the eleventh international conference on principles of knowledge representation and reasoning* (pp. 198-208). Menlo Park, CA: AAAI Press.)
- Halpern, J. Y. & Hitchcock, C. (2010). Actual causation and the art of modeling. (In R. Dechter, H. Geffner, & J. Y. Halpern (Eds.), *Heuristics, probability and causality: A tribute to Judea Pearl* (pp. 383-406). London: College Publications.)
- Halpern, J. Y. & Hitchcock, C. (ms.). Graded causation and defaults.
- Halpern, J. Y. & Pearl, J. (2001). Causes and explanations: A structural-model approach – Part I: Causes. (In J. Breese & D. Koller (Eds.), *Uncertainty in artificial intelligence: Proceedings of the seventeenth conference* (pp. 194–202). San Francisco: Morgan Kaufmann.)
- Halpern, J. Y. & Pearl, J. (2005). Causes and explanations: A structural-model approach. Part I: Causes. *British Journal for the Philosophy of Science*, 56, 843-887
- Hiddleston, E. (2005). Causal powers. *British Journal for the Philosophy of Science*, 56, 27-59
- Hitchcock, C. (2001). The intransitivity of causation revealed in equations and graphs. *Journal of Philosophy*, 98, 273-299
- Hitchcock, C. (2007a). Prevention, preemption, and the principle of sufficient reason.

⁷ Interestingly, David Danks, in his contribution to this special issue, argues that the concept of actual causation has a strong *prospective* aspect (connected to prediction and control), from which its *retrospective* aspect (connected to explanation) can't be divorced.

- Philosophical Review, 116, 495-532
- Hitchcock, C. (2007b). What's wrong with neuron diagrams? (In J. K. Campbell, M. O'Rourke, & H. S. Silverstein (Eds.), *Causation and explanation* (pp. 69-92). Cambridge, MA: MIT Press.)
- Hitchcock, C. (2009). Structural equations and causation: Six counterexamples. *Philosophical Studies*, 144, 391-401
- Hume, D. (1739-40). *A treatise of human nature*. (London: John Noon)
- Kvart, I. (2004). Causation: Probabilistic and counterfactual analyses. (In J. Collins, N. Hall & L. A. Paul (Eds.), *Causation and counterfactuals* (pp. 359-386). Cambridge, MA: MIT Press.)
- Lewis, D. (1973). Causation. *Journal of Philosophy*, 70, 556-567
- Lewis, D. (1979). Counterfactual dependence and time's arrow. *Noûs*, 13, 455-476
- Lewis, D. (1980). Veridical hallucination and prosthetic vision. *Australasian Journal of Philosophy*, 58, 239-249
- Lewis, D. (1986) Postscripts to "Causation". (In D. Lewis, *Philosophical papers*, Vol. II (pp. 172-213). Oxford: Oxford University Press.)
- Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science*, 67, 1-25
- Mackie, J. L. (1965). Causes and conditions. *American Philosophical Quarterly*, 2, 245-264
- Menzies, P. (2004). Causal models, token causation, and processes. *Philosophy of Science*, 71, 820-832
- Mill, J. S. (1843). *A system of logic, ratiocinative and inductive*, Vol. I. (London: John W. Parker)
- Mumford, S. & Anjum, R. L. (2011). *Getting causes from powers*. (Oxford: Oxford University Press)
- Pearl, J. (2000). *Causality: Models, reasoning and inference*. (Cambridge: Cambridge University Press). Second Edition published in 2009.
- Ramachandran, M. (2004). A counterfactual analysis of indeterministic causation. (In J. Collins, N. Hall & L. A. Paul (Eds.), *Causation and counterfactuals* (pp. 387-402). Cambridge, MA: MIT Press.)
- Spirtes, P., Glymour, C., & Scheines, R. (1993). *Causation, prediction, and search*. (Cambridge, MA: MIT Press). Second edition published in 2000.
- Strevens, M. (2007). Mackie remixed. (In J. K. Campbell, M. O'Rourke, & H. S. Silverstein (Eds.), *Causation and explanation* (pp. 93-118). Cambridge, MA: MIT Press.)
- Woodward, J. (2003). *Making things happen: A theory of causal explanation*. (New York: OUP)
- Wright, R. W. (1985). Causation in tort law. *California Law Review*, 73, 1735-1828
- Yablo, S. (2004). Advertisement for a sketch of an outline of a prototheory of causation. (In J. Collins, N. Hall & L. A. Paul (Eds.), *Causation and counterfactuals* (pp. 119-137). Cambridge, MA: MIT Press.)