Diagrammatic Definitions of Causal Claims

McHugh, D.

DOI
10.1007/978-3-319-91376-6_32

Publication date
2018

Document Version
Author accepted manuscript

Published in
Diagrammatic Representation and Inference

Diagrammatic Definitions of Causal Claims*

Dean McHugh

Institute of Logic, Language and Computation (ILLC),
University of Amsterdam

February 2017

Abstract

We present a class of diagrams in which to reason about causation. These diagrams are based on a formal semantics called 'system semantics', in which states of systems are related according to temporal succession. Arguing from straightforward examples, we provide the truth conditions for causal claims that one may make about these diagrams.

1 Introduction

Diagrams offer a natural and highly expressive means of depicting causal relations. Flowcharts are the ubiquitous example, but even more concerted work to analyse causal relations specifically employs an abundance of visual aids. Lewis (1973, 564), for instance, diagrammatically depicts similarity orderings over worlds, while Spirtes et al. (2000) and Pearl (2009) represent Bayes nets as directed acyclic graphs.

In this paper we follow the diagrammatic tradition by presenting a class of diagrams in which to reason about causation. The bulk of the work consists in presenting a variety of cases in which diagrams represent causal relations. Regarding the underlying formal apparatus, we construct these diagrams from a semantics called 'system semantics'. In Section 2 we outline the approach of system semantics and see how it may be used to characterise causal relations. Section 3 provides an alternative, diagrammatic characterisation of

*This is a preprint of an article published in the proceedings of *Diagrams: 10th International Conference on the Theory and Application of Diagrams*, June 18th-22nd 2018. The final authenticated version is available online at: https://doi.org/10.1007/978-3-319-91376-6_32.

these causal relations, and Section 4 refines the account by depicting two notions of 'sometimes' and 'partial' causation. In Section 5 we consider some further expressive power of diagrams in system semantics, showing how they may represent an agent's interaction with a system, and in Section 6 we conclude by outlining avenues for future work.

2 System Semantics for Causal Claims

To begin by analogy, system semantics aims to do for causal claims what Kripke semantics has achieved in the philosophical discussion of possible and necessary truth. Indeed, we modify Kripke semantics for modal logic to create diagrams called 'systems' that specify precisely how parts of possible worlds change through time. A system $\mathcal{S}$ is a pair $\langle St, R\rangle$ composed of a set St of states and a relation R of temporal succession between them. Each state represents a moment type rather than token, and is formally a valuation of atomic sentences in propositional logic. Given states s and t, the intuitive reading of "s is related to t" is that, if s is the current state, t may be the next state after one step in time.

To illustrate, suppose we have two atomic sentences representing a switch being up (S), and a light being illuminated (L). Figure 1 represents the interaction between the switch and light. Circles depict states, accompanied by the sentences that are true at them, and arrows depict the succession relation. The diagram of Figure 1 shows, for instance, that when the switch is up and light is off ($S, \neg L$) the system changes into the state where the switch is up and the light is on (S, L). And the top-left loop demonstrates that if the switch is down and light is off ($\neg S, \neg L$), then they remain so in the next state.

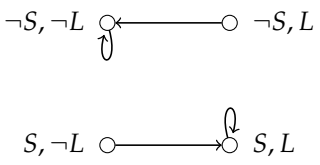

Figure 1: System composed of a switch and light.

Looking at Figure 1, it seems the following causal claim should come out true.

The switch being up is a necessary and sufficient cause of the light being on.

The truth conditions that we propose here for such a causal claim draw on the notion of a state's *past* and *future* states. These are encoded by a system's

relation of temporal succession R as the states leading to and from each state along R. We say that the switch being up is a necessary cause of the light being on because, in every state where the light is on, we see the switch was up in the past. And the switch being up is a sufficient cause of the light being on because every state where the switch is up leads only to states where the light is on.

In general, of course, we also have to require that the above claims are not trivially satisfied, as would happen, say, if the system featured only states where the switch is up and the light is on. Triviality would result as well if the light never changed, in the sense that states where the light is off lead only to states where the light is off, and states where the light is on lead only to states where the light is on. In a slogan, then, we must additionally require that the switch being up makes some *difference* to the light being on.

We can deal with these worries of triviality by providing the minimal conditions that a causal relation must satisfy. Let us say that a state *s leads to* a state *t*—and conversely, *t comes from s*—just in case there is some path along R from s to t. We then define the following notion of 'minimal causation'.

Definition 1 (Minimal cause)**.** *A is a minimal cause of B just in case*

(1a) Some B-state comes from or leads to some ¬B-state,
(1b) Some A-state leads to some B state, and
(1c) Some ¬A-state leads to some ¬B-state.

With Definition 1 providing the minimal conditions that a causal relation must satisfy, from the point of view of system semantics, we strengthen the conditions to define the notions of necessary and sufficient causation. To do so, let us say that state *s must lead to* state *t* just in case *t* eventually occurs from *s*, no matter what path the system takes from *s*. Likewise, *t must come from s* just in case *s* always occurred prior to *t*, no matter what path the system took to *t*.

We then strengthen the notion of minimal causation like so.

Definition 2 (Necessary cause)**.** *A is a necessary cause of B just in case*

(2a) A is a minimal cause of B, and
(2b) every B-state must come from some A-state.

Definition 3 (Sufficient cause)**.** *A is a sufficient cause of B just in case*

(3a) A is a minimal cause of B, and
(3b) every A-state must lead to some B-state.

Condition (2b) expresses that whenever B currently holds, A must have held at some point in the past, no matter what path the system took to the current state. Condition (3b) expresses that whenever A currently holds, B will hold at some point in the future, no matter what path from the current state the system will take. The reader is invited to check that, according to Definitions 2 and 3, in the system of Figure 1, S is indeed a necessary and sufficient cause of L.

It turns out that Definitions (1)–(3) above can be displayed in a purely diagrammatic way, as the next section demonstrates.

3 A Diagrammatic Definition

Given a system $\mathcal{S}$, we diagrammatically represent condition (1a) by saying that the system in question must feature some path depicted in Figure 2. For example, a system $\mathcal{S}$ features the topmost arrow from Figure 2 just in case some state of $\mathcal{S}$ where both A and B are false leads to a state where A is false and B true. (For convenience we suppress '$\neg A$' and '$\neg B$' in Figures 2 and 3.)

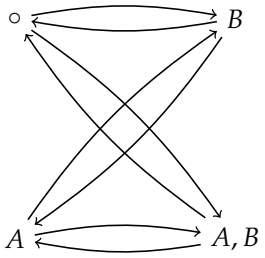

Figure 2: Some B-state leads to some $\neg B$-state, or vice versa.

We likewise represent conditions (1b) and (1c) by means of the unshaded diagrams appearing in Figure 3. That is, some A-state leads to some B-state in a system $\mathcal{S}$—i.e. (1b) holds—just in case $\mathcal{S}$ features some path from the bottom-right diagram of Figure 3. And some $\neg A$-state leads to some $\neg B$-state—i.e. (1c) holds—just in case $\mathcal{S}$ features some path from the top-left diagram of Figure 3.

The shaded diagrams of Figure 3 correspond to the definitions of necessary and sufficient causation, where this time we read its arrows in terms of the 'must' mode of coming and leading. That is, a system $\mathcal{S}$ satisfies condition (2b) just in case $\mathcal{S}$ features no path from the bottom-left diagram, while condition (3b) holds in a system $\mathcal{S}$ just in case $\mathcal{S}$ features no path from the top-right diagram.

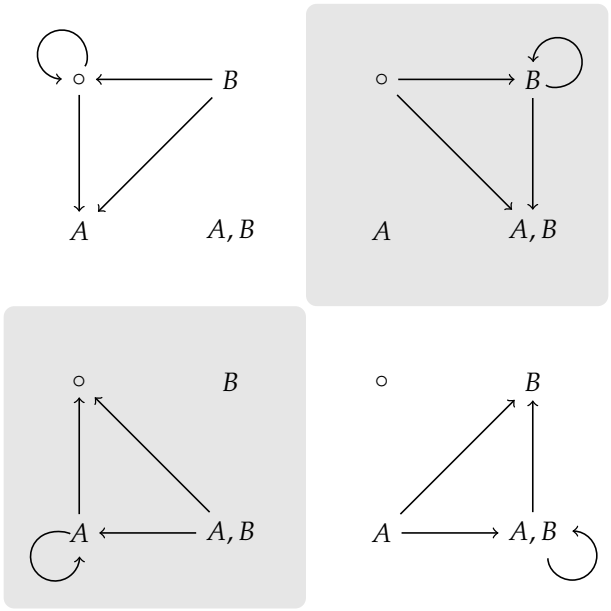

Figure 3: Diagrams depicting paths from states to states.

The definition of minimal causation given above is too weak on its own to serve as a definition of any intuitive notion of cause. For, conditions (1a)–(1c) only demand *some* paths of some specified kind, and so are even satisfied in systems in which states succeed one another in a completely random fashion; that is, in which every state leads to all states. In contrast, the definitions of necessary and sufficient causation are each more stringent by demanding that some paths are excluded from the system.

But one might also wonder whether they are too strong to adequately capture our causal talk. There seem to be many shades of causation falling short of conditions (2b) and (3b) that our diagrams should hope to represent. In the next section we consider two less demanding ways that a causal relation—such as minimal, necessary and sufficient causation—may hold in a system. These weaker modes of causation we call 'sometimes' and 'partial' causation.

4 Sometimes and Partial Causation

The purpose of introducing a 'sometimes' modifier into causal relations is to capture causal reasoning in non-deterministic systems. Now, many analyses of causation assume that the phenomena they wish to model behave deter-

ministically. We will not pursue the matter here, but only point out that two of the most popular analyses of causation assume some form of determinism. Firstly, Lewis's counterfactual analysis presumes a notion of determinism in order to account for the asymmetry of causal dependence (see Menzies, 2017, §2.2). Secondly, as Cartwright (1999) notes, the Bayes nets approach of Pearl (2009) and Spirtes et al. (2000) assumes determinism in order to satisfy one of their key assumptions, known as the Causal Markov Condition.

There are, nonetheless, many everyday processes we wish to model in which causes do not uniquely determine their effects. Consider, for instance, a computer with a faulty 'on' button, where pushing the button only sometimes succeeds in turning the computer on. (Or, more extremely, imagine the button's success depends on some quantum set up.) This on its own is a perfectly intelligible scenario, but analyses of causation that assume determinism can only model it by introducing extraneous variables; say, by introducing a hidden variable representing the button successfully connecting with the computer. System semantics avoid such complication by allowing states to have multiple successors. Thus, in system semantics we can straightforwardly depict this scenario by means of the diagram of Figure 4. We assume that the act of pushing the button lasts only one moment; that is, if the button is pushed at a state, then it reverts to being unpushed in the next state.

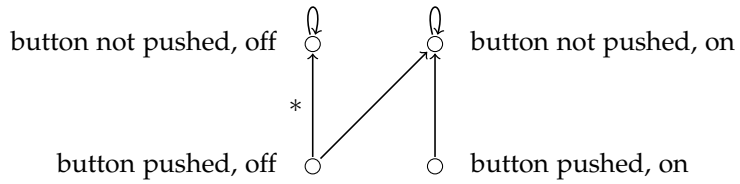

Figure 4: Pushing the button sometimes causes the computer to turn on.

The non-deterministic behaviour of the button corresponds to the fact that there are two arrows coming from the state where the button is pushed and the computer is off. If the system is in that state (pushed, off), then *sometimes*—i.e. when the button did not work and the system moved along the arrow marked with a star—the computer is still off in the next state. But at other times, when the button happens to work, in the next state the computer is on.

In some cases we wish to explicitly add extra variables into our models. This occurs, say, when we want to make a background assumption explicit, or reveal the influence of a previously hidden variable. Thus, for instance, one can take into account the presence or absence of charge in the computer of Figure 4: when there is charge (C), the system behaves as in Figure 4, but

when there is no charge ($\neg C$) the system always moves into a state where the computer is off. Figure 5 illustrates this new system.

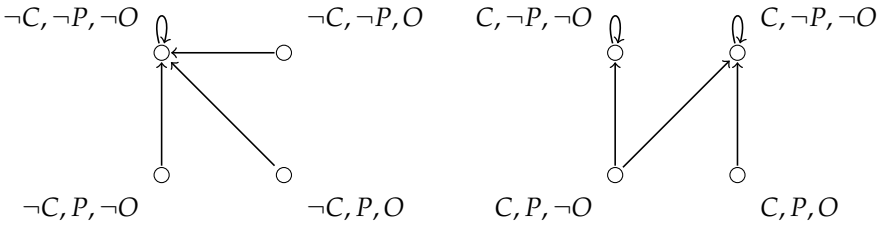

Figure 5: When there is charge (C), pushing the button (P) sometimes causes the computer to turn on (O).

Upon examination of Figures 4 and 5, it seems reasonable to assert the following causal claims.

(4a) In the system of Figure 4, pushing the button is sometimes a sufficient cause of the computer turning on.

(4b) In the system of Figure 5, when there is charge, pushing the button is sometimes a sufficient cause of the computer turning on.

In Figure 4 we see that the path responsible for introducing the qualification 'sometimes' into (4a) is the path marked with a star. It is because of this arrow that not every path from a pushed state leads to the computer being on, meaning the system of Figure 4 does not satisfy condition (3b). Hence, according to Definition 3, pushing the button is not a sufficient cause of the computer being on. Nonetheless, were we to restrict attention to just those times when pushing the button *is* successful—by removing the contravening arrow from the diagram—then pushing the button would be a sufficient cause of the computer turning on. This suggests the following truth condition for adding a 'sometimes' modifier to a given causal relation, defined by means of operations on diagrams.

Definition 4 (Sometimes relation)**.** *A causal relation holds sometimes, in a system $\mathcal{S}$, just in case it holds by removing some (possibly no) arrows from $\mathcal{S}$.*

The system of Figure 4 makes (4a) true since, in the system that results from removing the arrow marked with a star, pushing the button is a sufficient cause of the computer turning on.

Turning now to Figure 5, it seems we want to say that pushing the button sometimes causes the computer to turn on, but only when there is charge.

We can give the truth conditions for such conditional assertions by taking up an idea of Kratzer (1991), whereby conditionals are restrictions on quantifiers. In the present context, the proposal amounts to saying that a statement of the form 'If A then B' is true just in case B is true with respect to the A-states. Such a notion of conditional causal claims we call a notion of 'partial' causation, because for the causal claim to hold it need only hold in *part* of the model.

Definition 5 (Partial relation)**.** *A causal relation holds partially, in a system $\mathcal{S}$, just in case it holds by removing some (possibly no) states from $\mathcal{S}$.*

Note that the definitions of sometimes and partial causation above imply that every partial relation is also a sometimes relation. For we can mimic the result of removing states from a system by removing every arrow that touches a state we wish to remove. But we cannot go the other way: there are sometimes relations that are not partial relations, as happens whenever we have to remove some but not all arrows touching a given state. This occurs, for instance, in the system of Figure 4 because removing any state where the computer is off—which is enough to remove the arrow marked with a star—would also make the system falsify condition (1c) and fail even the test for minimal causation.

A further advantage of depicting causal relations in terms of system semantics is that one may naturally consider multiple relations holding in the same diagram. The following section briefly outlines how such a proposal can be used to model an agent's interaction with a system.

5 Modelling an Agent's Interaction

By focusing on changes of states individually, rather than the global behaviour of variables, system semantics provides a novel level of detail absent from other approaches, notably the structural causal models of Halpern (2000) and Pearl (2009, §7.1). One advantage of the finer grain of system semantics is the abundance of ways to represent relations between states. For example, as some have demanded of automata (e.g. Baeten et al., 2011), we may naturally add succession relations to represent different kinds of change—such as those brought about by a user interacting with a system and those brought about by the system itself.

Figures 6a and 6b depict two different ways to add an interaction relation to the system depicted in Figure 1. The dark lines indicate changes made by the system independently (nature's path, so to speak), while the dashed lines

depict a user's path, interacting with the system. In Figure 6a turning off the switch immediately turns off the light, whereas Figure 6b the user takes a turn, only after which the system reacts.

(a) Interaction with immediate effect

(b) Interaction with delayed effect

Figure 6: Two ways to add interaction to a system.

Extrapolating from this simple scenario, we may model multi-agent games by introducing one relation for each agent over states of a gameplay.

6 Conclusion

In this paper we saw some simple diagrams depict the modelling power of system semantics. Of course, one must invest quite some work just to provide a system-semantic representation of any given process, prior to analysing its causal relations. In this brief exposition we have made no argument for the capacity of the diagrams of system semantics to represent every kind of process we would wish to model. But given the widespread use of causal notions in diverse fields, such an argument would be required if system semantics for causal claims is to properly fulfil its representational ambition.

We further saw how, by encoding temporal succession into the models directly, we could analyse causal notions in a fairly straightforward manner. Of course, we have not touched upon the metaphysical issues underlying such an approach; for instance, we took the notion of temporal succession to be unproblematic. A more comprehensive appraisal of system semantics must examine whether the choices of primitives made by system semantics fare better than those of other approaches to causality, such as the assumption of a similarity ordering over worlds made by Lewis (1973). One benefit of system semantics is that its metaphysical commitment—chiefly, an ontology of states related in time—is reasonably transparent, though to fully make the case for the philosophical adequacy of system semantics, one must still argue that those are sensible commitments to make.

References

Jos CM Baeten, Bas Luttik, and Paul van Tilburg. Computations and interaction. *ICDCIT*, 6536:35–54, 2011. doi:10.1007/978-3-642-19056-8_3.

Nancy Cartwright. Causal diversity and the markov condition. *Synthese*, 121 (1):3–27, 1999. doi:10.1023/A:1005225629681.

Joseph Y Halpern. Axiomatizing causal reasoning. *Journal of Artificial Intelligence Research*, 12:317–337, 2000.

Angelika Kratzer. Conditionals. In Arnim von Stechow and Dieter Wunderlich, editors, *Semantics: An international handbook of contemporary research*, pages 639–650. Berlin:de Gruyter, 1991.

David Lewis. Causation. *Journal of Philosophy*, 70(17):556–567, 1973. doi:10.2307/2025310.

Peter Menzies. Counterfactual theories of causation. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2017 edition, 2017. URL `https://plato.stanford.edu/archives/win2017/entries/causation-counterfactual/`.

Judea Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, New York, NY, USA, 2nd edition, 2009.

Peter Spirtes, Clark N Glymour, and Richard Scheines. *Causation, prediction, and search*. MIT press, 2000.