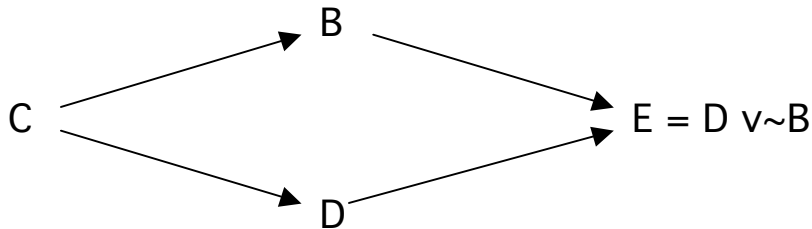


**Sufficient reason.** Problem for the causal network approach. I call it the “faux hero” problem, after the firefighter who is there to put the fire out only because she started it. Suppose  $c$  the my placing the bomb under your chair;  $d$  is your fleeing upon spotting the bomb;  $b$  is the chair’s blowing up;  $e$  is your surviving. Here’s the graph, annotated to clarify the dependence relations:



The path from C to E through B is inactive, since holding off-path variable D fixed at its actual value 1, E is 1 regardless of what value C takes. But the path through D is active, since holding B fixed at its actual value 1, E is 1 iff C = 1. So the account wrongly predicts that my faux-heroic act of putting a bomb under your chair caused your survival, by tipping you off to the impending explosion. How should the network approach deal with this?

*Unintuitive.* “The interpolated variable B is not easy to find, and the ...counterfactual that reveals the active causal route from [C to E] is not at all intuitive. [For] the relevant piece of counterfactual reasoning would go as follows: Suppose that the [chair had blown up], and suppose moreover that that [there had been no bomb under it] in the first place. Since [there was no bomb, you] would not have seen [the explosion] coming and would not have [run away]; it would have [occurred right under you] and [you] would not have survived.... This is the sort of counterfactual reasoning that only trained philosophers engage in; unaided intuition is not to be faulted for failing to “see” the relevant...counterfactual.” *Reply.* This seems less like a refutation of the imagined trained philosopher than an apologia on behalf of the poor non-philosopher. But the non-philosopher, if she denies  $c$  causes  $e$ , is correct!

*Far-fetched.* The above figure “does not constitute an appropriate representation” of [BOMB]...our causal judgments depend on which unactualized possibilities we are willing to take seriously, and which we consider too remote. The variables we choose to include should reflect our concerns...When we exclude the variable B from our model...[that is because] we are not willing to take seriously the possibility that the [chair explodes even though there is no bomb under it]. This possibility is just too far fetched.” *Reply.* Change the example a bit; C is my pushing a delicately balanced bomb off a ledge to fall on your chair; D is your noticing my doing this and fleeing; B is the chair’s exploding; E is your survival. It’s not at all plausible that my action was a cause of your survival. But it’s not at all far-fetched that the delicately balanced bomb would fall for some other reason.

*Sufficient reason (1).* Distinguish a variable  $V$ 's taking a *default* value from its taking a *deviant* value. Intuitively, default values are what you'd expect anyway; deviant values depart from the norm in a way that intuitively requires explanation. The *principle of sufficient reason* says that a variable  $V$  with parents  $V_1 \dots V_n$  does not take a deviant value unless one of the  $V_i$ s takes a deviant value. Call a network *self-contained* if every variable in the network satisfies the principle of sufficient reason. "The main idea I will defend [called TC for token causation] is that counterfactual dependence [dependence not holding anything fixed] is necessary and sufficient for token causation in self-contained networks."

*Example.*  $C$  = my placing the bomb.  $D$  = your fleeing.  $E$  is your surviving.  $D = C$  and  $E = D \vee \sim C$ . This network is self-contained (why?). So we look for straight counterfactual dependence, not holding anything fixed. There is none:  $E = 1$  if  $C = 1$  because then  $D = 1$ , and  $E = 1$  if  $C = 0$  because  $\sim C$  is a disjunct of  $E$ . So we now have a story about why the faux hero is not a cause.

*Complaint.* The network account's pride and joy is standard early preemption cases, e.g., Suzy throwing so that Billy doesn't bother. But the causal network here is not self-contained, for  $B$  takes a deviant value (1 for Billy-throw) when its lone parent  $S$  assumes its default value (0 for Suzy-no-throw). This means some of our strongest positive judgments are no longer confirmed by the theory.

*Sufficient reason (2).* "The reader cannot help but have noticed that TC provides a necessary condition for causation, and a sufficient condition, but leaves a gap between them. What are we to say about cases where the causal network is not self-contained? ...What I suspect is that something like the [active-route account] is essentially correct, but that the verdicts of such an account can be partially or completely counteracted by TC when the latter yields a clear negative conclusion." The suggestion then seems to be this:  $c$  causes  $e$  iff EITHER the network is self-contained and  $e$  depends on  $c$  absolutely, that is holding nothing fixed (note sufficient reason is automatically satisfied), or it is not self-contained and  $e$  depends on  $c$  holding off-path variables fixed at their default values (so as to minimize violations of sufficient reason).

*Reply.* Suppose Bodyguard inserts antidote ( $D = 1$ ) iff he sees Assassin inserting poison ( $A = 1$ ). Then Victim survives ( $E = 1$ ) no matter what. Intuitively what Bodyguard does saves Victim's life; it's a cause of Victim's survival. But the network is self-contained (why?). And we saw there is no counterfactual dependence; Victim survives no matter what. Conversely we can change BOMB so the network is not self-contained; the bomb is dropped because I remove the note telling my assistant *not* to drop it. Still seems like faux heroism, but now we're told causation goes with an active path, and there is one; holding fixed that the chair was going to explode, you survive iff you are tipped off by my removing the note.