

The Transitivity and Asymmetry of Actual Causation^{*}

Sander Beckers¹ and Joost Vennekens²

¹ Cornell University

² KU Leuven, Dept. Computer Science @ Campus De Nayer

The problem of actual causation has received a lot of attention in both the philosophy and the AI literature. In a nutshell, the problem is to define when event X is deemed to have caused event Y in the context of a particular story. Typically, it is assumed that this story unfolds according to a given set of causal laws. Coming up with a suitable definition for this concept of actual causation has proven to be quite difficult.

In this paper, we attempt to both explain why this is so difficult, and to offer a method for potentially solving the problem. We do so by focusing on two natural, yet mutually incompatible intuitions, namely that actual causation should be both transitive and asymmetric.

Following the approach of the seminal work by Halpern and Pearl [2], we make use of structural models as our formal tool.

Example 1. An assassin-in-training is on his first mission. Trainee is an excellent shot: if he shoots his gun, the bullet will fell Victim. Supervisor is also present, in case Trainee has a last minute loss of nerve (a common affliction among student assassins) and fails to pull the trigger. If Trainee does not shoot, Supervisor will shoot Victim herself. In fact, Trainee performs admirably, firing his gun and killing Victim.

We can represent this example by means of a structural model that consists of five boolean random variables and two equations that relate these variables:

$$\begin{aligned} VictimDies &:= TraineeHits \vee SupervisorHits \\ SupervisorHits &:= SupervisorShoots \\ TraineeHits &:= TraineeShoots \\ SupervisorShoots &:= \neg TraineeShoots \end{aligned}$$

The story told in the example corresponds to the following assignment of values to these variables:

$$\begin{aligned} TraineeShoots &= TraineeHits = VictimDies = true; \\ SupervisorShoots &= SupervisorHits = false. \end{aligned}$$

In his seminal work, Lewis [3] proposed to define actual causation as the transitive closure of counterfactual dependency. In the case of the above example,

^{*} The full version of this paper was published as [1].

VictimDies is counterfactually dependent on *TraineeHits* (in the context of the story, in which *SupervisorHits* is false, it is the case that if trainee’s bullet hadn’t hit the victim, the victim would not have died) and *TraineeHits* is counterfactually dependent on *TraineeShoots* (if he hadn’t shot, then his bullet wouldn’t have hit the victim). Therefore, by transitivity, victim’s death was caused by trainee’s shooting, even though it does not counterfactually depend on it (if trainee would not have shot, then victim would still have died).

The transitivity of actual causation invoked by Lewis seems intuitively plausible. Indeed, if X caused Y and Y in turn caused Z , then surely it is fair to call also X a cause of Z . However, several convincing counterexamples have emerged, such as the following example by [4].

Example 2. Terrorist, who is right-handed, must push a detonator button at noon to set off a bomb. Shortly before noon, he is bitten by a dog on his right hand. Unable to use his right hand, he pushes the detonator with his left hand at noon. The bomb duly explodes.

$$Bomb := LeftHand \vee RightHand$$

$$LeftHand := DogBite$$

$$RightHand := \neg DogBite$$

Clearly, the dog bite caused the terrorist to use his left hand, and the terrorist’s use of his left hand caused the explosion. Nevertheless, it seems too farfetched to call the dog bite a cause for the explosion, as transitivity would mandate. Therefore, this example demonstrates that the intuition of transitivity has its limitations.

In this paper, we claim that the limiting factor is the intuition of asymmetry, i.e., that we should never consider X to have caused Y if, at the same time, we would also have considered $\neg X$ (if $\neg X$ had been the case instead of X) as a cause of the same Y . In the case of the above example, we see that, in the context where the dog would not have bitten the terrorist, there would be counterfactual dependence of *Bomb* on *RightHand* and of *RightHand* on $\neg DogBite$. Therefore, Lewis’ account would count both *DogBite* and $\neg DogBite$ as a cause for the same event *Bomb*, which would violate asymmetry.

In the full version of this paper, we argue for the hypothesis that asymmetry is the limiting factor on the transitivity of actual causation, i.e., that actual causation is precisely as transitive as it can be without violating asymmetry.

References

1. Sander Beckers and Joost Vennekens. The transitivity and asymmetry of actual causation. *Ergo, an Open Access Journal of Philosophy*, 2017.
2. J. Halpern and J. Pearl. Causes and explanations: A structural-model approach. part i: Causes. *The British Journal for the Philosophy of Science*, 56:843–87, 2005.
3. D. Lewis. Causation. *Journal of Philosophy*, 70:113–126, 1973.
4. M. McDermott. Redundant causation. *The British Journal for the Philosophy of Science*, 46(4):523–544, 1995.