A Theory of Actual Causation

Lennart Ackermans

Munich Institute for Mathematical Philosophy

August 28, 2025

What is a Theory of Actual Causation?

- Consider any scenario with two actually occurring events.

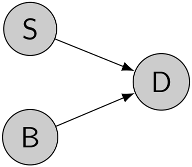
Scenario

Suzy and Billy throw a rock at an empty bottle. The rocks hit the bottle at the same time and it breaks.

- A theory of actual causation is a simple algorithm that tells you whether one event caused another.
- The challenge: find a theory that aligns with the causal intuitions of professional philosophers, computer scientists and other academics.

Scenario

Suzy (S) and Billy (B) throw a rock at an empty bottle. The rocks hit the bottle at the same time and it is destroyed (D). Either rock is sufficient for the rock breaking.



Equations:

$D \doteq S \vee B$

Valuation:

$S \doteq 1$

$B \doteq 1$

Definition (Intervention)

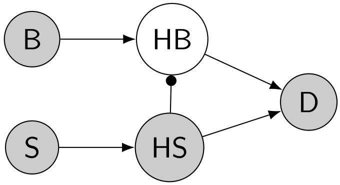
An *intervention* on $X \doteq x$ sets X to x and removes the equation for X.

The Need for Better Theories

- Almost all existing theories violate some intuitions, including Halpern (2015), Hall (2007), Beckers (2021), Halpern and Hitchcock (2015), and Gallow (2021).
- Andreas and Günther (forthcoming) seems to satisfy all intuitions.
- My theory makes the same judgments as Andreas and Günther (forthcoming), but it is simple, intuitive and counterfactual.

Scenario

Billy (B) and Suzy (S) throw rocks at a bottle. Suzy's rock hits the bottle first (HS), shattering it (D). If Suzy's rock had not hit the bottle, then Billy's rock would have hit it (HB) and shattered it.



Equations:

$HB \doteq B \wedge \neg HS$

$HS \doteq S$

$D \doteq HB \vee HS$

Valuation:

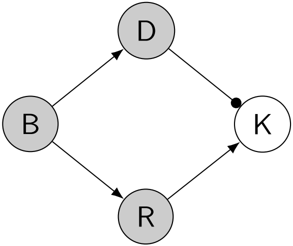
$B \doteq 1$

$S \doteq 1$

LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Scenario

A boulder is dislodged (B) and rolls towards a hiker (R). The hiker sees the boulder dislodgement and ducks (D). As a result, the hiker survives. If the hiker had not ducked, he would have died (K).



Equations:

$D \doteq B$

$R \doteq B$

$K \doteq R \wedge \neg D$

Valuation:

$B \doteq 1$

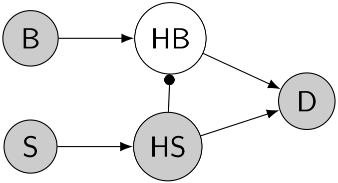
Definition (Model-Relative Causation)

$A \doteq a$ causes $B \doteq b$ relative to a model M if and only if

1. $A \doteq a$ and $B \doteq b$ occur in M.
2. There exists an intervention set I_1 which
   - intervenes on $A \doteq a$, and
   - may intervene on non-actual values of exogenous variables, and
   - may intervene on actual values of endogenous variables that are not a descendant of A.

   and an intervention set I_2 which
   - intervenes on $A \doteq \neg a$, and
   - intervenes on the new value of every variable changed as a result of the intervention on I_1,

   such that
   (a) $B \doteq b$ after intervening on I_1,
   (b) $B \doteq \neg b$ after intervening on both I_1 and I_2.



Equations:

$HB \doteq B \land \neg HS$

$HS \doteq S$

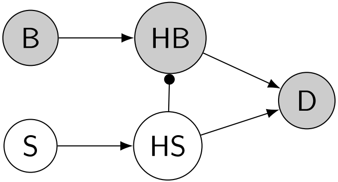
$D \doteq HB \lor HS$

Valuation:

$B \doteq 1$

$S \doteq 1$

- Let $I_1 = \{B \doteq 1, S \doteq 0\}$. This yields:



Equations:

$HB \doteq B \land \neg HS$

$HS \doteq S$

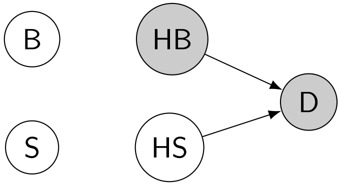
$D \doteq HB \lor HS$

Valuation:

$B \doteq 1$

$S \doteq 0$

- HB and HS have changed in value, so we have $I_2 = \{B \doteq 0, HB \doteq 1, HS \doteq 0\}$.
- Intervening on $I_2 = \{B \doteq 0, HB \doteq 1, HS \doteq 0\}$ yields:



Equations:

$HB \doteq B \wedge \neg HS$

$HS \doteq S$

$D \doteq HB \vee HS$

Valuation:

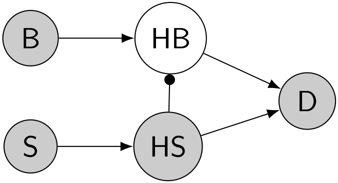
$B \doteq 0$

$S \doteq 0$

$HB \doteq 1$

$HS \doteq 0$

- In this model D is true, so B does not cause D.



Equations:

$HB \doteq B \wedge \neg HS$

$HS \doteq S$

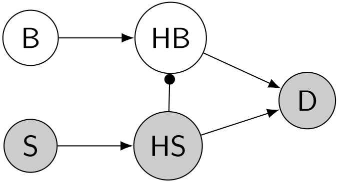
$D \doteq HB \vee HS$

Valuation:

$B \doteq 1$

$S \doteq 1$

- Let $I_1 = \{S \doteq 1, B \doteq 0\}$. Intervening yields:



Equations:

$HB \doteq B \wedge \neg HS$

$HS \doteq S$

$D \doteq HB \vee HS$

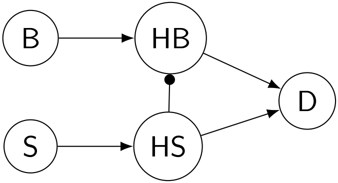
Valuation:

$B \doteq 0$

$S \doteq 1$

- Nothing has changed in value, so we have $I_2 = \{S \doteq 0\}$.

- Intervening on $I_2 = \{S \doteq 0\}$ yields:



Equations:

$HB \doteq B \wedge \neg HS$

$HS \doteq S$

$D \doteq HB \vee HS$

Valuation:

$B \doteq 0$

$S \doteq 0$

- In this model D is false, so S causes D.

Some restrictions on model appropriateness

- A model M is appropriate for some scenario only if:
  - **Exogenous deviancy.** For every exogenous variable X, its actual value $X \doteq x$ is more deviant than $X \doteq \neg x$.
  - **Exogenous saturation.** One cannot add a new exogenous variable X to M such that $M + X$ is appropriate (in other senses).
  - **Endogenous saturation 1.** One cannot add a new endogenous variable X to M, such that X has at least two parents or two children and $M + X$ is appropriate (in other senses).
  - **Endogenous saturation 2.** When a variable A in M has more than one parent or child and C, with no or two or more children, is a descendant of A, then M must have a variable B on the path between A and C with exactly one parent and child.
  - (Apart from the above, appropriateness requires accuracy and a number of other conditions.)

A definition of causation

Definition (Actual Causation 1)

An event **a** causes an event **b** in some scenario **S** if and only if **there exists** an appropriate model M such that $A \doteq a$ causes $B \doteq b$ with respect to M.

Definition (Actual Causation 2)

An event **a** causes an event **b** in some scenario **S** if and only if **for all** appropriate models M, $A \doteq a$ causes $B \doteq b$ with respect to M.

Model Invariance

Conjecture (Model-Invariance)

Let **S** be a scenario and let M_1, M_2 be appropriate models for **S**. Then **a** causes **b** with respect to M_1 if and only if **a** causes **b** with respect to M_2.

Theorem (Local Model-Invariance)

Let **S** *be a scenario and let M an appropriate model for* **S**. *Let $M + X$ be an appropriate model for* **S**, *where the interpretation for the variables in M remains the same. Then* **a** *causes* **b** *in M if and only if* **a** *causes* **b** *in $M + X$.*

Achievements of the theory

- My theory gives the intuitively correct judgments on the following types of scenario (and more):
  - Simple overdetermination
  - Late preemption
  - Early preemption
  - Threat and saviour (short circuit)
  - Double prevention
  - Preempted prevention
  - Switches
  - Bogus prevention
- My theory is simpler and more intuitive than similarly successful theories.

Andreas, Holger, and Mario Günther. Forthcoming. "Factual Difference-Making." *Australasian Philosophical Review.*

Beckers, Sander. 2021. "Causal Sufficiency and Actual Causation." *Journal of Philosophical Logic* 50, no. 6 (December): 1341–1374.

Gallow, J. Dmitri. 2021. "A Model-Invariant Theory of Causation." *The Philosophical Review* 130, no. 1 (January): 45–96.

Hall, N. 2007. "Structural equations and causation." *Philosophical Studies* 132, no. 1 (January): 109–136.

Halpern, Joseph Y. 2015. *A Modification of the Halpern-Pearl Definition of Causality.*

Halpern, Joseph Y., and Christopher Hitchcock. 2015. "Graded Causation and Defaults." *The British Journal for the Philosophy of Science* 66 (2): 413–457.

Download slides