Actual Causation

Causal Models and Meanings – Class IV

Dean McHugh

University of Edinburgh
Department of Linguistics

Yerevan Academic for Linguistics and Philosophy
YALP 2026

Plan

1. Halpern's *Actual Causality* (2016)
   - Halpern's definition of actual causation
   - Minimality
2. Applying Halpern's analysis: discrimination law
3. Halpern's account of overdetermination
   - The Billy and Suzy case
   - Ned Hall's switching scenario

Plan

1. Halpern's *Actual Causality* (2016)
   - Halpern's definition of actual causation
   - Minimality
2. Applying Halpern's analysis: discrimination law
3. Halpern's account of overdetermination
   - The Billy and Suzy case
   - Ned Hall's switching scenario

Plan

1. Halpern's *Actual Causality* (2016)
   - Halpern's definition of actual causation
   - Minimality
2. Applying Halpern's analysis: discrimination law
3. Halpern's account of overdetermination
   - The Billy and Suzy case
   - Ned Hall's switching scenario

Halpern (2016)

Definition (Truth conditions for actual causal claims)

$X = x$ is an *actual cause* of φ in the causal setting (M, u) iff

AC1 $(M, u) \models X = x$ and $(M, u) \models \varphi$

Halpern (2016)

Definition (Truth conditions for actual causal claims)

$X = x$ is an *actual cause* of φ in the causal setting (M, u) iff

AC1 $(M, u) \models X = x$ and $(M, u) \models \varphi$

AC2 There is a set W of variables and a setting x' of the variables in X such that $(M, u) \models W = w$ and $(M, u) \models [X \leftarrow x', W \leftarrow w]\neg\varphi$

Halpern (2016)

Definition (Truth conditions for actual causal claims)

$X = x$ is an *actual cause* of φ in the causal setting (M, u) iff

AC1 $(M, u) \models X = x$ and $(M, u) \models \varphi$

AC2 There is a set W of variables and a setting x' of the variables in X such that $(M, u) \models W = w$ and $(M, u) \models [X \leftarrow x', W \leftarrow w]\neg\varphi$

AC3 X is minimal; there is no strict subset X' of X such that $X' = x'$ satisfies conditions AC1 and AC2, where x' is the restriction of x to the variables in X'

Halpern (2016)

Definition (Truth conditions for actual causal claims)

$X = x$ is an *actual cause* of φ in the causal setting (M, u) iff

AC1 $(M, u) \models X = x$ and $(M, u) \models \varphi$

AC2 There is a set W of variables and a setting x' of the variables in X such that $(M, u) \models W = w$ and $(M, u) \models [X \leftarrow x', W \leftarrow w]\neg\varphi$

AC3 X is minimal; there is no strict subset X' of X such that $X' = x'$ satisfies conditions AC1 and AC2, where x' is the restriction of x to the variables in X'

AC2 Intuition: "we can show the counterfactual dependence of φ on X by keeping the variables in W fixed at their actual values" (Halpern, 2016, p. 25)

Halpern (2016)

Definition (Truth conditions for actual causal claims)

$X = x$ is an *actual cause* of φ in the causal setting (M, u) iff

AC1 $(M, u) \models X = x$ and $(M, u) \models \varphi$

AC2 There is a set W of variables and a setting x' of the variables in X such that $(M, u) \models W = w$ and $(M, u) \models [X \leftarrow x', W \leftarrow w]\neg\varphi$

AC3 X is minimal; there is no strict subset X' of X such that $X' = x'$ satisfies conditions AC1 and AC2, where x' is the restriction of x to the variables in X'

AC2 Intuition: "we can show the counterfactual dependence of φ on X by keeping the variables in W fixed at their actual values" (Halpern, 2016, p. 25)

Halpern (2016)

Definition (Truth conditions for actual causal claims)

$X = x$ is an *actual cause* of φ in the causal setting (M, u) iff

AC1 $(M, u) \models X = x$ and $(M, u) \models \varphi$

AC2 There is a set W of variables and a setting x' of the variables in X such that $(M, u) \models W = w$ and $(M, u) \models [X \leftarrow x', W \leftarrow w]\neg\varphi$

AC3 X is minimal; there is no strict subset X' of X such that $X' = x'$ satisfies conditions AC1 and AC2, where x' is the restriction of x to the variables in X'

AC2 Intuition: "we can show the counterfactual dependence of φ on X by keeping the variables in W fixed at their actual values" (Halpern, 2016, p. 25)

Questions:

1. What if W is empty?
2. What if we pick a W such that $(M, u) \not\models W = w$?

Plan

1. Halpern's *Actual Causality* (2016)
   - Halpern's definition of actual causation
   - Minimality
2. Applying Halpern's analysis: discrimination law
3. Halpern's account of overdetermination
   - The Billy and Suzy case
   - Ned Hall's switching scenario

Minimality

Definition (Truth conditions for actual causal claims)

$X = x$ is an *actual cause* of φ in the causal setting (M, u) iff

AC1 $(M, u) \models X = x$ and $(M, u) \models \varphi$

AC2 There is a set W of variables and a setting x' of the variables in X such that $(M, u) \models W = w$ and $(M, u) \models [X \leftarrow x', W \leftarrow w]\neg\varphi$

AC3 X is minimal; there is no strict subset X' of X such that $X' = x'$ satisfies conditions AC1 and AC2, where x' is the restriction of x to the variables in X'

Example

Company proposals pass just in case both committee members vote in favour. A proposal came along. Both committee members voted in favour, so it passed.

(1) The fact that both committee members voted in favour of the proposal caused it to pass.

Minimality

Example

Company proposals pass just in case both committee members vote in favour. A proposal came along. Both committee members voted in favour, so it passed.

(2) The fact that both committee members voted in favour of the proposal caused it to pass.

$$A = 1 \wedge S = 1 \text{ caused } P = 1$$

Minimality

Example

Company proposals pass just in case both committee members vote in favour. A proposal came along. Both committee members voted in favour, so it passed.

(2) The fact that both committee members voted in favour of the proposal caused it to pass.

$A = 1 \wedge S = 1$ caused $P = 1$

Without AC3, if dropping a lit match qualified as a cause of the forest fire, then dropping a match and sneezing would also pass the tests of AC1 and AC2. AC3 serves here to strip "sneezing" and other irrelevant, over-specific details from the cause.

(Halpern, 2016, 23)

Plan

1. Halpern's *Actual Causality* (2016)
   - Halpern's definition of actual causation
   - Minimality
2. Applying Halpern's analysis: discrimination law
3. Halpern's account of overdetermination
   - The Billy and Suzy case
   - Ned Hall's switching scenario

L.A. Department of Water & Power v. Manhart (1978)

- L.A. Department of Water & Power required female employees to pay higher pension contributions.
- The reason: on average, women live longer than men, and therefore require a larger pension.
- Marie Manhart filed a class action suit of behalf of female employees of L.A. Water & Power.

Syllabus available at `https://www.loc.gov/item/usrep435702/`

L.A. Department of Water & Power v. Manhart (1978)

- L.A. Department of Water & Power required female employees to pay higher pension contributions.
- The reason: on average, women live longer than men, and therefore require a larger pension.
- Marie Manhart filed a class action suit of behalf of female employees of L.A. Water & Power.

Syllabus available at `https://www.loc.gov/item/usrep435702/`

- Let's work with this analysis of discrimination:

  x was discriminated against on the basis of trait t just in case x having trait t caused x to be harmed.

For an insightful criticism of this approach, see Kohler-Hausmann (2018).

Manhart in detail

- Employees receive the same annual pension from age 65 to when they die.
- Suppose Marie Manhart died at age 80.
- If she had been a man, she would have paid less into the fund.
- Since she died at age 80, she received 15 years' worth of pension money.
- Let x's *net contribution* = amount paid − amount received.

The counterargument

L.A. Department of Water & Power says:

> *"This case has nothing to do with sex and everything to with life expectancy. If Marie Manhart had been a man, yes, she would have paid less into the fund, but she probably would have lived longer and so taken more money out. We calculated the difference to ensure that, on average, the net contribution of men and women is the same."*

Your challenge

- Draw a causal model of the Manhart case.
- Determine whether, on Halpern's analysis, Marie Manhart being a woman cause her to pay a higher net contribution.

Your challenge

- Draw a causal model of the Manhart case.
- Determine whether, on Halpern's analysis, Marie Manhart being a woman cause her to pay a higher net contribution.

Hint. Try using these five variables:

1. Age of death
2. Amount paid into fund
3. Amount received from fund
4. Net contribution
5. Sex

Justice Stevens, *Manhart* Opinion

Myths and purely habitual assumptions about a woman's inability to perform certain kinds of work are no longer acceptable reasons for refusing to employ qualified individuals, or for paying them less. This case does not, however, involve a fictional difference between men and women. It involves a generalization that the parties accept as unquestionably true: Women, as a class, do live longer than men.

The question, therefore, is whether the existence or nonexistence of "discrimination" is to be determined by comparison of class characteristics or individual characteristics.

Source: `https://www.loc.gov/item/usrep435702/`

Justice Stevens, *Manhart* Opinion

If height is required for a job, a tall woman may not be refused employment merely because, on the average, women are too short. Even a true generalization about the class is an insufficient reason for disqualifying an individual to whom the generalization does not apply.

[T]here is no assurance that any individual woman working for the Department will actually fit the generalization on which the Department's policy is based. Many of those individuals will not live as long as the average man. While they were working, those individuals received smaller paychecks because of their sex, but they will receive no compensating advantage when they retire

Such a practice does not pass the simple test of whether the evidence shows "treatment of a person in a manner which but for that person's sex would be different."

Source: `https://www.loc.gov/item/usrep435702/`.

Plan

1. Halpern's *Actual Causality* (2016)
   - Halpern's definition of actual causation
   - Minimality
2. Applying Halpern's analysis: discrimination law
3. Halpern's account of overdetermination
   - The Billy and Suzy case
   - Ned Hall's switching scenario

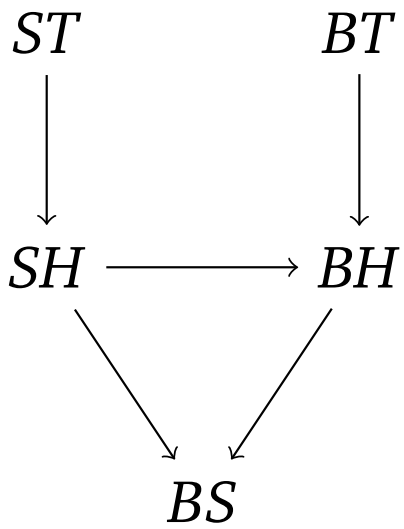
Plan

1. Halpern's *Actual Causality* (2016)
   - Halpern's definition of actual causation
   - Minimality
2. Applying Halpern's analysis: discrimination law
3. Halpern's account of overdetermination
   - The Billy and Suzy case
   - Ned Hall's switching scenario

The Billy and Suzy case

Example (Late preemption)

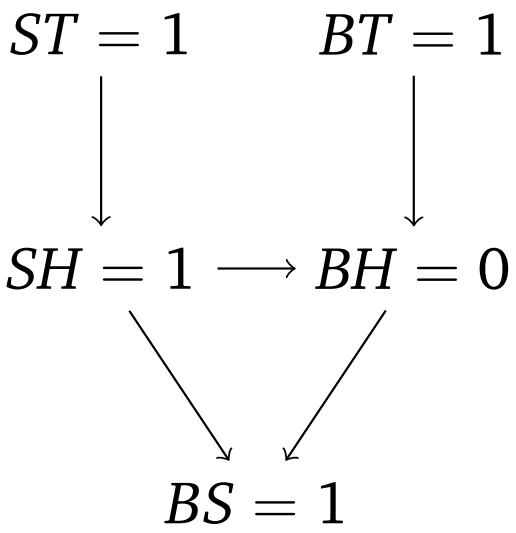
Suzy and Billy both throw a rock at a bottle. Suzy's rock gets there first, shattering the bottle. However Billy's throw was also accurate, and would have shattered the bottle had it not been preempted by Suzy's throw.



$$SH = ST$$
$$BH = BT \wedge \neg SH$$
$$BS = SH \vee BH$$

Figure: Halpern's model of *Late preemption* (2016, p. 31)

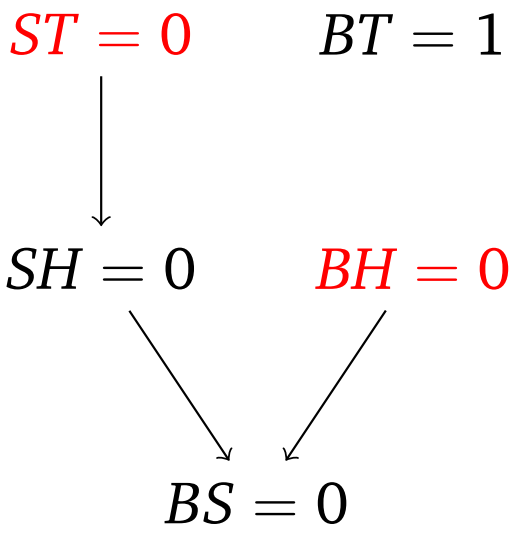
Halpern's account of the Billy and Suzy case



$$SH = ST$$
$$BH = BT \wedge \neg SH$$
$$BS = SH \vee BH$$

Figure: Halpern's model of *Late preemption* (2016, p. 31)

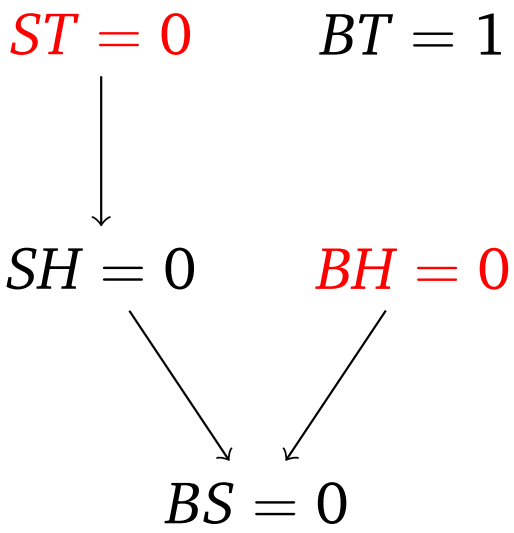
Halpern's account of the Billy and Suzy case



$$SH = ST$$
$$BH = 0$$
$$BS = SH \vee BH$$

Figure: Halpern's model of *Late preemption* (2016, p. 31)

Halpern's account of the Billy and Suzy case



$$SH = ST$$
$$BH = 0$$
$$BS = SH \vee BH$$

Figure: Halpern's model of *Late preemption* (2016, p. 31)

$ST = 1$ actually caused $BS = 1$ in (M, u). Take $W = BH$

AC1 $M, u \models ST = 1$ and $M, u \models BS = 1$.

AC2 $M, u \models BH = 1$ and $M, u \models [ST = 0, BH = 1]\neg(BS = 1)$.

AC3 ST is minimal.

Plan

1. Halpern's *Actual Causality* (2016)
   - Halpern's definition of actual causation
   - Minimality
2. Applying Halpern's analysis: discrimination law
3. Halpern's account of overdetermination
   - The Billy and Suzy case
   - Ned Hall's switching scenario

Ned Hall's switching scenario

A train rushes toward a fork in the tracks. If a switch is flipped, the train will take the left track, and if the switch is left in its original position, the train will take the right track. Further on, the left and the right tracks merge, and just after they meet, there is a train station.
Alex is standing at the switch when a train approaches. She flips the switch, so the train takes the left track. It rejoins the right track and reaches the station.

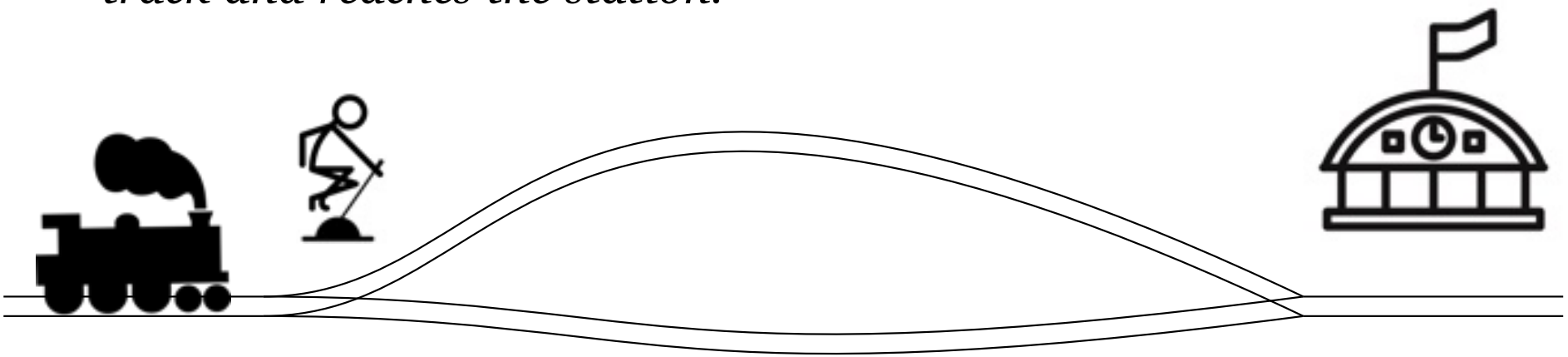


Figure: Switching scenario from Hall (2000, p. 205)

Ned Hall's switching scenario

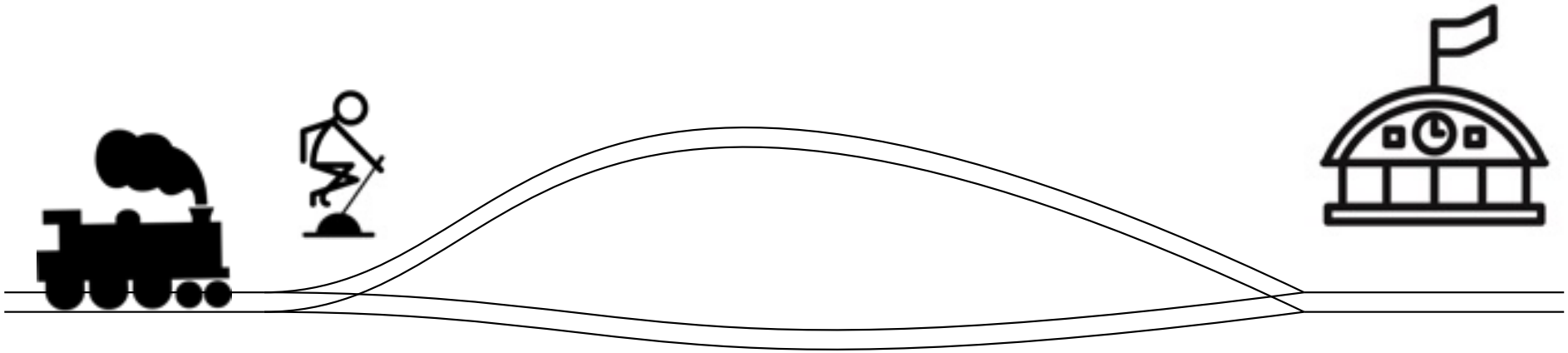
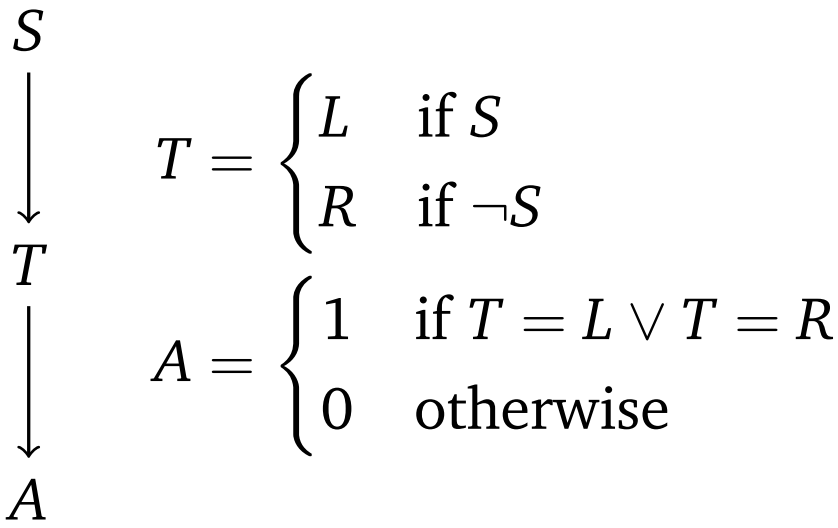


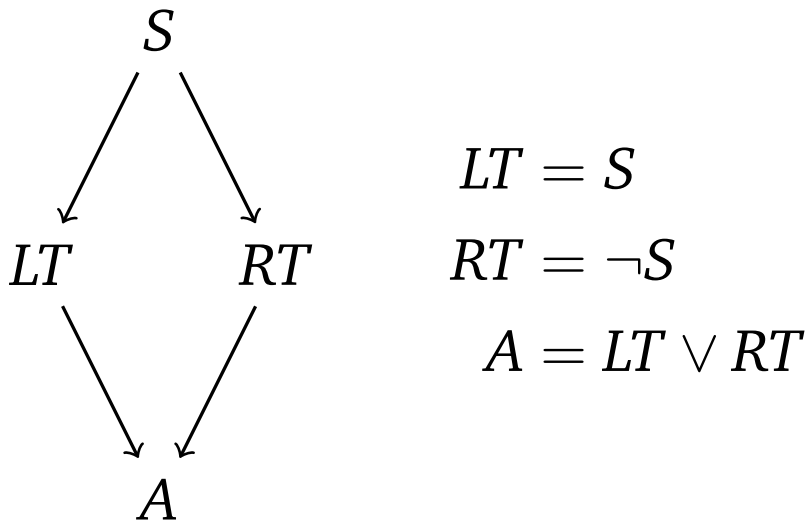
Figure: Switching scenario from Hall (2000, p. 205)

(3) Alex flipping the switch caused the train to reach the station.

Two models of the switching scenario



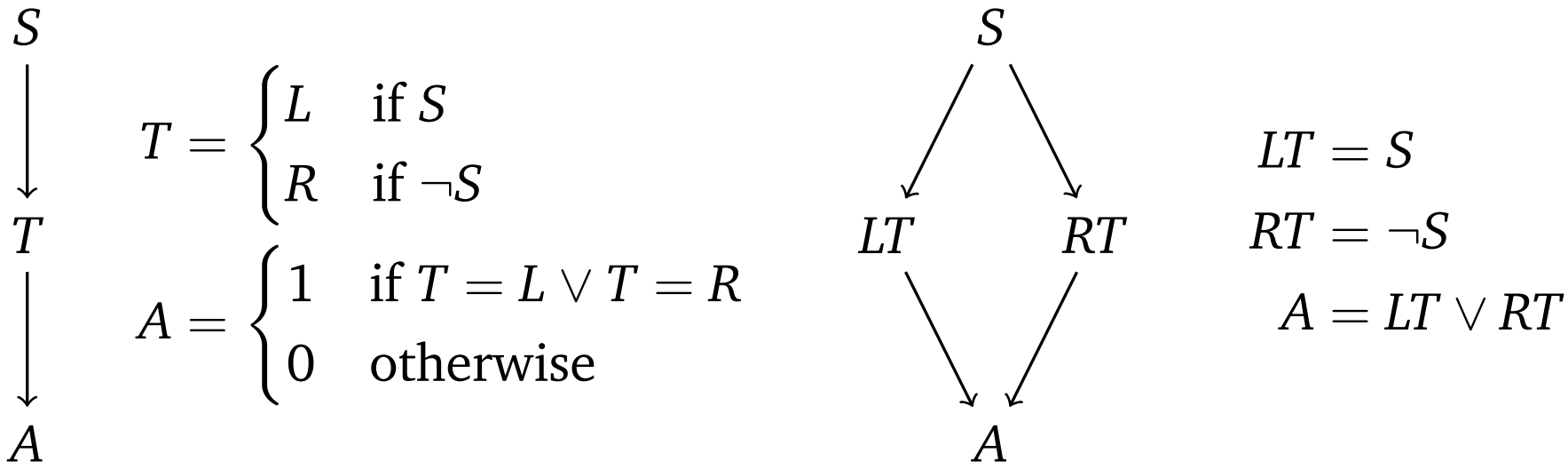
(a) One-variable model



(b) Two-variable model

Figure: Two models of the switching scenario

Two models of the switching scenario



(a) One-variable model

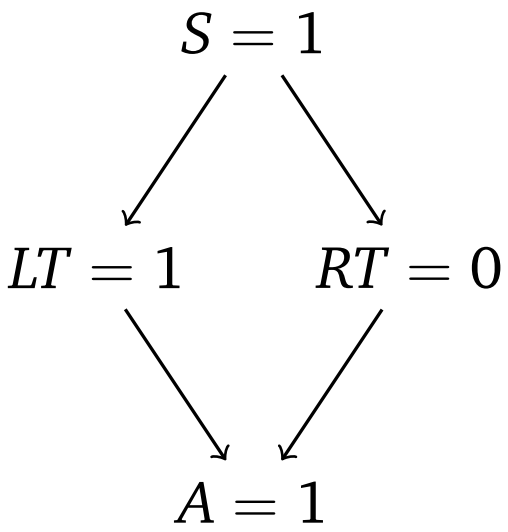
(b) Two-variable model

Figure: Two models of the switching scenario

(3) Alex flipping the switch caused the train to reach the station.

Question: What truth value does (3) have in each of the models above, according to Halpern (2016)?

The two-variable model



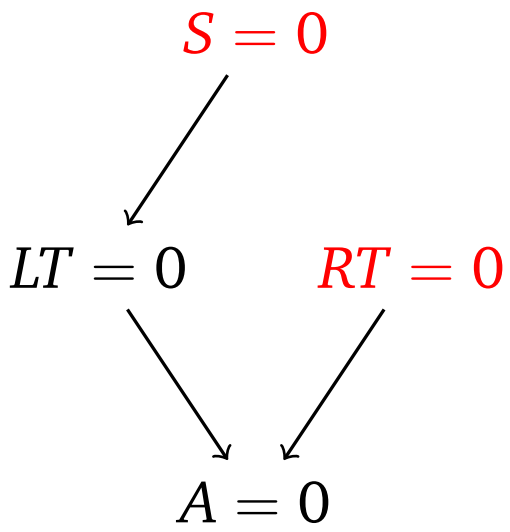
$$LT = S$$
$$RT = \neg S$$
$$A = LT \vee RT$$

Figure: Two-variable model

According to Halpern (2016), in the two-variable model, (3) is true.

(3) Alex flipping the switch caused the train to reach the station.

The two-variable model



$$LT = S$$
$$RT = 0$$
$$A = LT \vee RT$$

Figure: Two-variable model

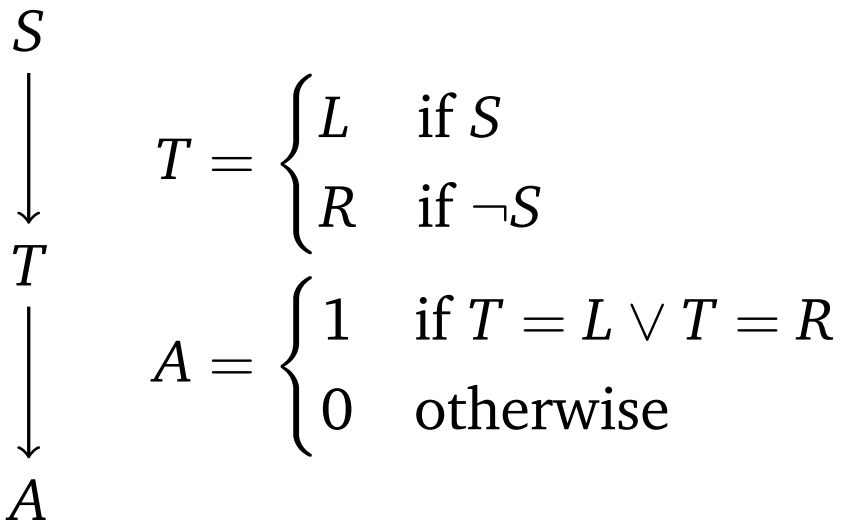
According to Halpern (2016), in the two-variable model, (3) is true.

(3) Alex flipping the switch caused the train to reach the station.

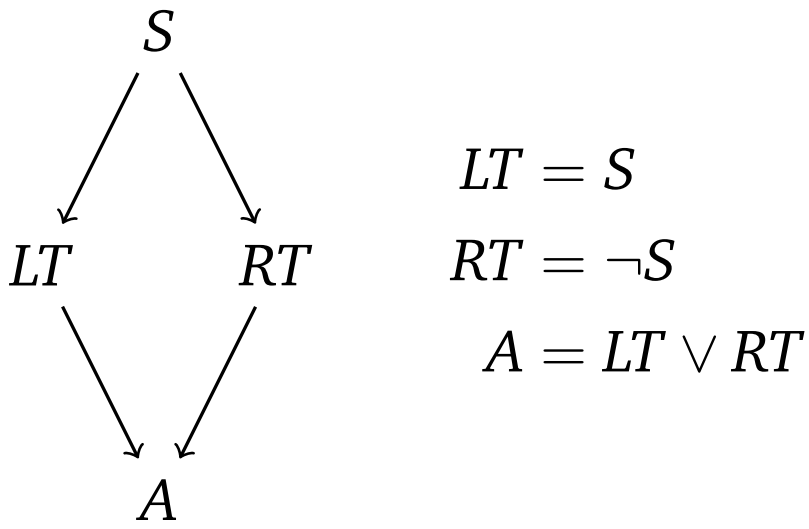
Comparing the two models, Halpern and Pearl (2005, p. 872) write:

> *The two-variable model depicts the tracks as two independent mechanisms, thus allowing one track to be set (by action or mishap) to false (or true) without affecting the other. Specifically, this permits the disastrous mishap of flipping the switch while the left track is malfunctioning. More formally, it allows a setting where* $S = 1$ *and* $RT = 0$*. Such abnormal settings are imaginable and expressible in the two-variable model, but not in the one-variable model.*

The two-variable model also allows a setting where $S = 0$ and $RT = 0$. The one-variable model rules this out as part of its variable structure.



(a) One-variable model



(b) Two-variable model

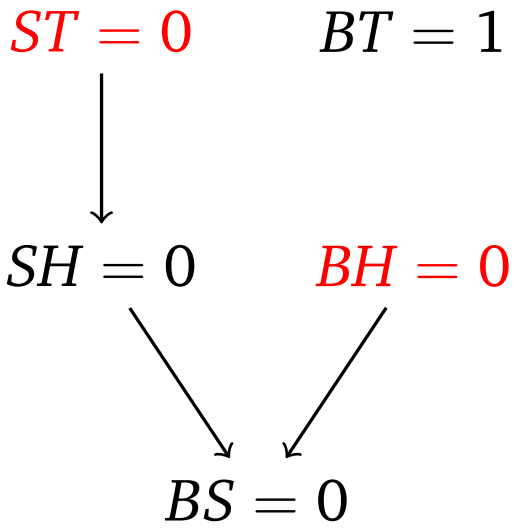
Figure: Two models of the switching scenario

In the two-variable model, one can intervene to make

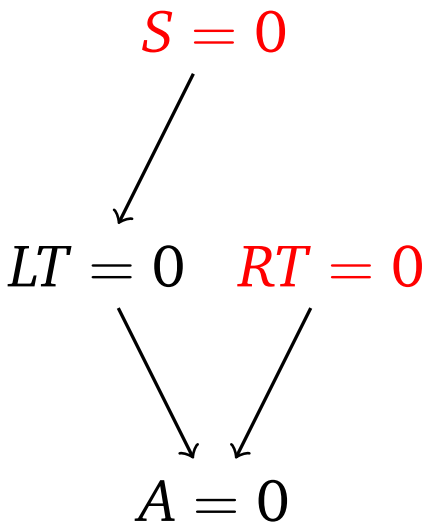
$$S = 0, LT = 0 \text{ and } RT = 0.$$

That is, interventions can make train disappear from the tracks!

The two-variable model



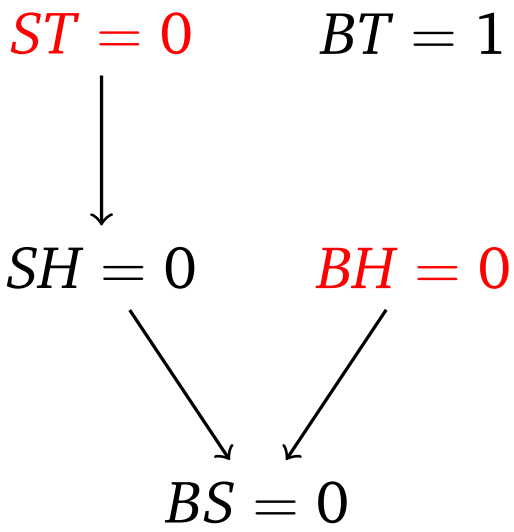
(a) Witness to Suzy causing the bottle to break



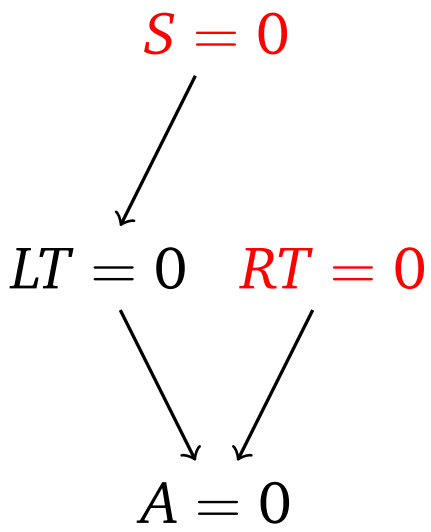
(b) Witness to the switch causing the train to arrive

Figure: Comparing explanations in overdetermination cases

The two-variable model



(a) Witness to Suzy causing the bottle to break

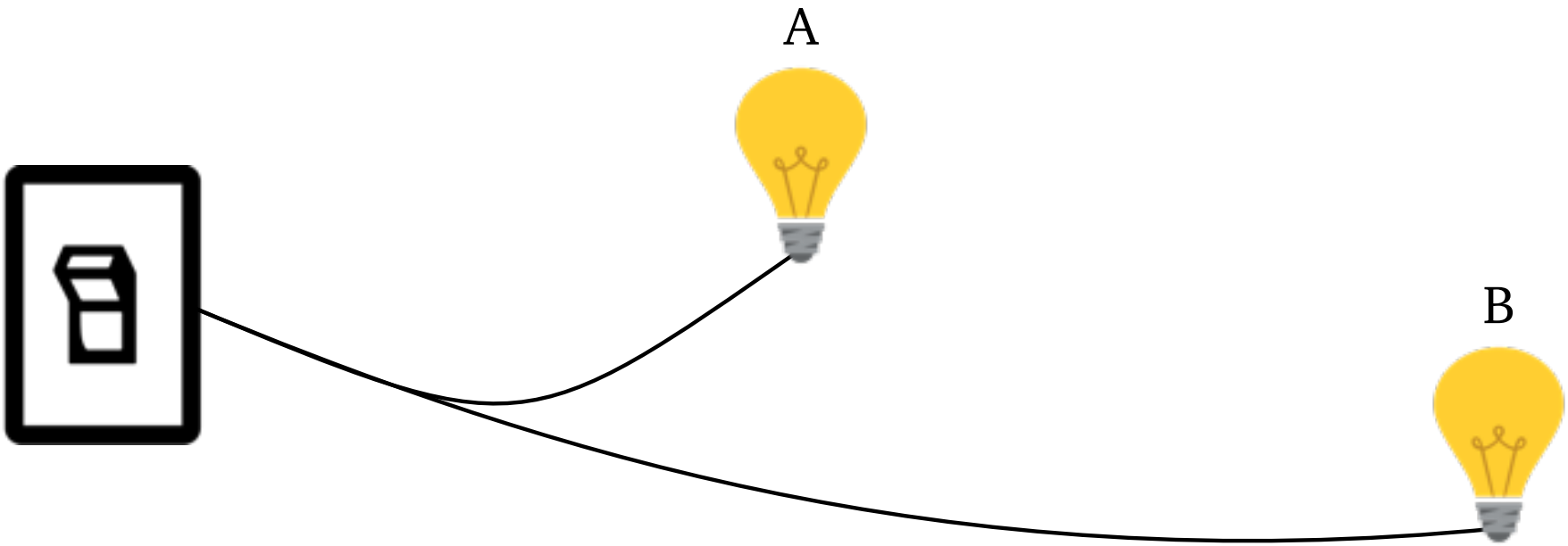


(b) Witness to the switch causing the train to arrive

Figure: Comparing explanations in overdetermination cases

If Billy's rock can disappear mid-flight,
why can't the train disappear mid-journey as well?

Comparison with the two lights case



$S = 1$

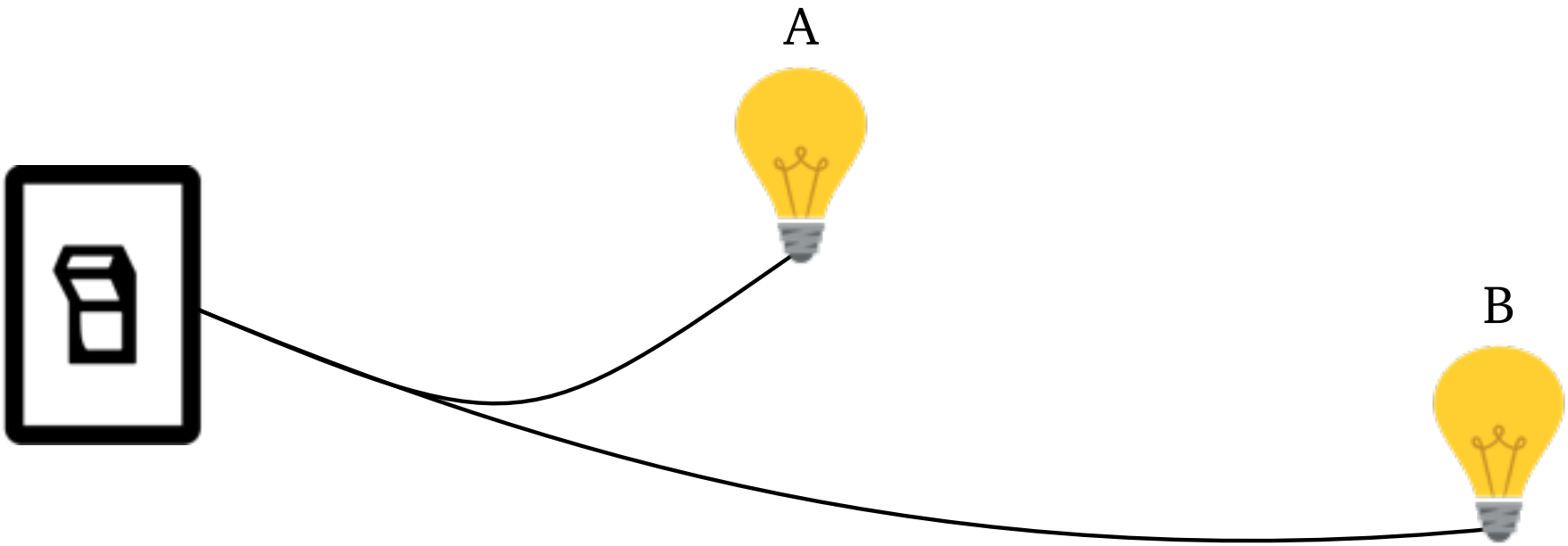
$A = 1$ $\quad$ $B = 1$

$A = S$

$B = S$

If the train can't disappear from the tracks,
why can electricity disappear from the wire?

Comparison with the two lights case



$S = 1$

$A = 0$ $\quad$ $B = 1$

$A = 0$

$B = S$

If the train can't disappear from the tracks,
why can electricity disappear from the wire?

The accessibility of violations in causal reasoning

- Halpern appeals to a violation to explain the judgement in the Billy and Suzy case: a scenario where Suzy does not throw her rock, Billy does, and still the bottle does not break
- The violation of the train disappearing does seem to not enter our reasoning with the switch scenario
- Halpern claims that a violation does enter our reasoning with the Billy and Suzy case
- Which violations are allowed?

Figure: Is Halpern opening a Pandora's box?

References I

Ned Hall. Causation and the price of transitivity. *Journal of Philosophy*, 97(4):198, 2000. doi:10.2307/2678390.

Joseph Y Halpern. *Actual Causality*. MIT Press, 2016.

Joseph Y Halpern and Judea Pearl. Causes and explanations: A structural-model approach. part i: Causes. *The British journal for the philosophy of science*, 56(4): 843–887, 2005.

Issa Kohler-Hausmann. Eddie murphy and the dangers of counterfactual causal thinking about detecting racial discrimination. *Nw. UL Rev.*, 113:1163, 2018. doi:10.2139/ssrn.3050650.