Unraveling Sartorio's Difference-Making Principle

Dean McHugh

APA (Eastern Division)
New York

Friday, 10 January 2025

INSTITUTE FOR LOGIC,
LANGUAGE AND COMPUTATION

UNIVERSITY
OF AMSTERDAM

(1) a. The employer did not hire Elisabeth Dekker **because** she is
         pregnant.
     b. ChatGPT being trained on far-right Reddit posts **caused** it to
         output racist stereotypes.

(1) a. The employer did not hire Elisabeth Dekker **because** she is
         pregnant.
     b. ChatGPT being trained on far-right Reddit posts **caused** it to
         output racist stereotypes.

The modelling question. What information do we use when we judge that
                a causal claim holds? In other words, what information
                should a causal model contain?

(1)   a.   The employer did not hire Elisabeth Dekker **because** she is
           pregnant.
      b.   ChatGPT being trained on far-right Reddit posts **caused** it to
           output racist stereotypes.

The modelling question. What information do we use when we judge that
                         a causal claim holds? In other words, what information
                         should a causal model contain?

The meaning question. Under what conditions is a causal claim true or
                       false? That is, what do causal claims mean?

Plan

"We think of a cause as something that makes a difference, and the difference it makes must be a difference from what would have happened without it."
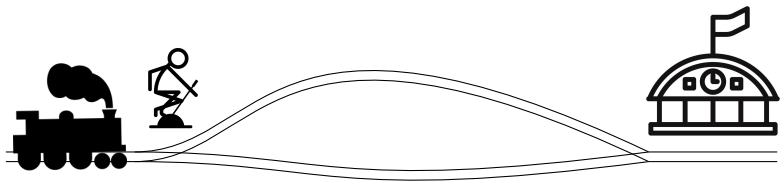
(Lewis 1973)

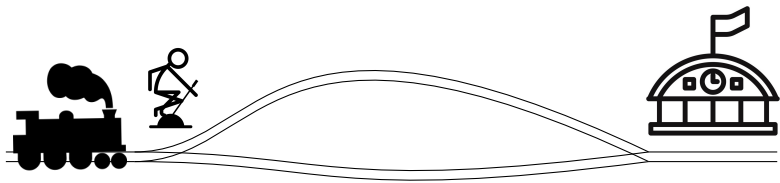Figure: Switching scenario from Hall (2000, p. 205).

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

(2)   a.   The train reached the station because the engineer flipped the switch.

       b.   The engineer flipping the switch caused the train to reach the station.

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

(3) a. The train reached the station because the engineer flipped the switch. ✗

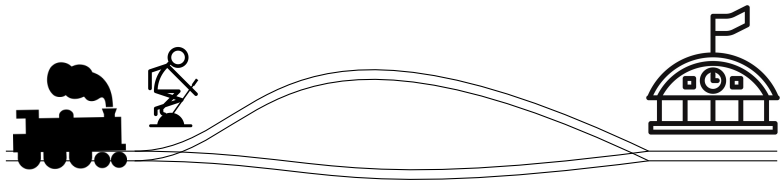b. The engineer flipping the switch caused the train to reach the station. ✗

C caused E
E because C

C caused E
E because C $\Rightarrow$ *C* made a difference to *E*

C caused E
E because C

$\Rightarrow$

C made a difference to *E*

Today's question

What does "*C* made a difference to *E*" mean?

Hypothesis: Difference making is counterfactual dependence

C made a difference to E

just in case

if C had not occurred, E would not have occurred.

It is not quite clear what 'dependence' is supposed to be, but at least it seems to imply that you would not get the effect without the cause.

The trouble with this is that you might from some other cause. That this effect was produced by this cause does not at all show that it could not, or would not, have been produced by something else in the absence of this cause.

(Anscombe 1971)

Suzy and Billy, expert rock-throwers, are engaged in a competition to see who can shatter a target window first.

They both pick up rocks and throw them at the window, but Suzy throws hers before Billy. Consequently, Suzy's rock gets there first, shattering the window.

Since both throws are perfectly accurate, Billy's would have shattered the window if Suzy's had not occurred.

(Hall and Paul 2003, p. 110; Hall 2004, p. 235)

Suzy and Billy, expert rock-throwers, are engaged in a competition to see who can shatter a target window first.

They both pick up rocks and throw them at the window, but Suzy throws hers before Billy. Consequently, Suzy's rock gets there first, shattering the window.

Since both throws are perfectly accurate, Billy's would have shattered the window if Suzy's had not occurred.

(Hall and Paul 2003, p. 110; Hall 2004, p. 235)

(4)   The window broke because Suzy threw her rock at it.

(5)   Suzy throwing her rock at the window caused it to break.

Suzy and Billy, expert rock-throwers, are engaged in a competition to see who can shatter a target window first.

They both pick up rocks and throw them at the window, but Suzy throws hers before Billy. Consequently, Suzy's rock gets there first, shattering the window.

Since both throws are perfectly accurate, Billy's would have shattered the window if Suzy's had not occurred.

(Hall and Paul 2003, p. 110; Hall 2004, p. 235)

(4)   The window broke because Suzy threw her rock at it.   ✓

(5)   Suzy throwing her rock at the window caused it to break.   ✓

If Suzy hadn't thrown her rock, the window would have broken anyway.

C caused *E* even though *E* does not counterfactually depend on *C*.

Hypothesis: Difference making is counterfactual dependence

C made a difference to E

just in case
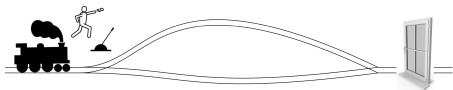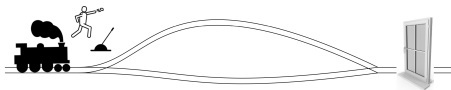
if C had not occurred, E would not have occurred.

The engineer pulling the lever
did not cause
the train to reach the station.

Suzy throwing her rock
did cause
the window to break.

Suzy throwing her rock
did not cause
the window to break.

Suzy throwing her rock
did cause
the window to break.

Plan

One thing that catches the eye … is that, just as the *flip* doesn't make a difference to the [train reaching the station], the *failure to flip* wouldn't have made a difference to the [train reaching the station] either. In other words, *whether or not* I flip the switch makes no difference [to the train's arrival], it only helps to determine the route that the train takes [to the station].

(Sartorio 2005, pp. 74–75)

Sartorio's Principle

If C caused E, then,
had C not occurred, the absence of C wouldn't have caused E.

Sartorio's Principle

If C caused E, then,
had C not occurred, the absence of C wouldn't have caused E.

$$C \text{ cause } E \quad \Rightarrow \quad \neg C > \neg(\neg C \text{ cause } E)$$

Sartorio's Principle accounts for the switch case:

- Suppose for reductio that the engineer pulling the lever caused the train to reach the station.

Sartorio's Principle accounts for the switch case:

- Suppose for reductio that the engineer pulling the lever caused the train to reach the station.
- Then (intuitively) if the engineer hadn't pulled the lever, that would have also caused the train to reach the station.

Sartorio's Principle accounts for the switch case:

- Suppose for reductio that the engineer pulling the lever caused the train to reach the station.
- Then (intuitively) if the engineer hadn't pulled the lever, that would have also caused the train to reach the station.
- Sartorio's principle is violated.

Sartorio's Principle accounts for the switch case:

- Suppose for reductio that the engineer pulling the lever caused the train to reach the station.
- Then (intuitively) if the engineer hadn't pulled the lever, that would have also caused the train to reach the station.
- Sartorio's principle is violated.
- Then assuming Sartorio's principle, the engineer pulling the lever did not cause the train to reach the station.

Sartorio's Principle also accounts for the Billy and Suzy case:

- Imagine that Suzy had not thrown.

Sartorio's Principle also accounts for the Billy and Suzy case:

- Imagine that Suzy had not thrown.
- In that case Billy's rock would have hit the window, and it would have broken anyway.

Sartorio's Principle also accounts for the Billy and Suzy case:

- Imagine that Suzy had not thrown.
- In that case Billy's rock would have hit the window, and it would have broken anyway.
- Intuitively, Billy's throw caused the window to break.

Sartorio's Principle also accounts for the Billy and Suzy case:

- Imagine that Suzy had not thrown.
- In that case Billy's rock would have hit the window, and it would have broken anyway.
- Intuitively, Billy's throw caused the window to break.
- But what about Suzy *not* throwing?

Sartorio's Principle also accounts for the Billy and Suzy case:

- Imagine that Suzy had not thrown.
- In that case Billy's rock would have hit the window, and it would have broken anyway.
- Intuitively, Billy's throw caused the window to break.
- But what about Suzy *not* throwing?

Sartorio's Principle also accounts for the Billy and Suzy case:

- Imagine that Suzy had not thrown.
- In that case Billy's rock would have hit the window, and it would have broken anyway.
- Intuitively, Billy's throw caused the window to break.
- But what about Suzy *not* throwing?

  Did that cause the window to break?

Sartorio's Principle also accounts for the Billy and Suzy case:

- Imagine that Suzy had not thrown.
- In that case Billy's rock would have hit the window, and it would have broken anyway.
- Intuitively, Billy's throw caused the window to break.
- But what about Suzy *not* throwing?

  Did that cause the window to break?

  Intuitively not!

Sartorio's Principle

If C caused E, then, had C not occurred, the absence of C wouldn't have caused E.

$$C \text{ cause } E \quad \Rightarrow \quad \neg C > \neg(\neg C \text{ cause } E)$$

Sartorio's Principle

If C caused E, then, had C not occurred, the absence of C wouldn't have caused E.

$$C \text{ cause } E \quad \Rightarrow \quad \neg C > \neg(\neg C \text{ cause } E)$$

Sartorio's Principle gives us a principled way to distinguish overdetermination and switching cases.

- Suzy's throw satisfies Sartorio's Principle.
- Pulling the lever does not.

If C caused E, then, had C not occurred, the absence of C wouldn't have caused E.

$$C \text{ cause } E \quad \Rightarrow \quad \neg C > \neg(\neg C \text{ cause } E)$$

A nice result: Sartorio's principle is automatically satisfied when the effect counterfactually depends on the cause.

- Suppose $\neg C > \neg E$

If C caused E, then, had C not occurred, the absence of C wouldn't have caused E.

$$C \text{ cause } E \quad \Rightarrow \quad \neg C > \neg(\neg C \text{ cause } E)$$

A nice result: Sartorio's principle is automatically satisfied when the effect counterfactually depends on the cause.

- Suppose $\neg C > \neg E$
- $\neg C$ cause E entails E,

Sartorio's Principle

If C caused E, then, had C not occurred, the absence of C wouldn't have caused E.

$$C \text{ cause } E \quad \Rightarrow \quad \neg C > \neg(\neg C \text{ cause } E)$$

A nice result: Sartorio's principle is automatically satisfied when the effect counterfactually depends on the cause.

- Suppose $\neg C > \neg E$
- $\neg C$ cause E entails E,
- By contraposition, $\neg E$ entails $\neg(\neg C \text{ cause } E)$

If C caused E, then, had C not occurred, the absence of C wouldn't have caused E.

$$C \text{ cause } E \quad \Rightarrow \quad \neg C > \neg(\neg C \text{ cause } E)$$

A nice result: Sartorio's principle is automatically satisfied when the effect counterfactually depends on the cause.

- Suppose $\neg C > \neg E$
- $\neg C$ cause E entails E,
- By contraposition, $\neg E$ entails $\neg(\neg C \text{ cause } E)$
- If X entails Y then $A > X$ entails $A > Y$

If C caused E, then, had C not occurred, the absence of C wouldn't have caused E.

$$C \text{ cause } E \quad \Rightarrow \quad \neg C > \neg(\neg C \text{ cause } E)$$

A nice result: Sartorio's principle is automatically satisfied when the effect counterfactually depends on the cause.

- Suppose $\neg C > \neg E$
- $\neg C$ cause E entails E,
- By contraposition, $\neg E$ entails $\neg(\neg C \text{ cause } E)$
- If X entails Y then $A > X$ entails $A > Y$
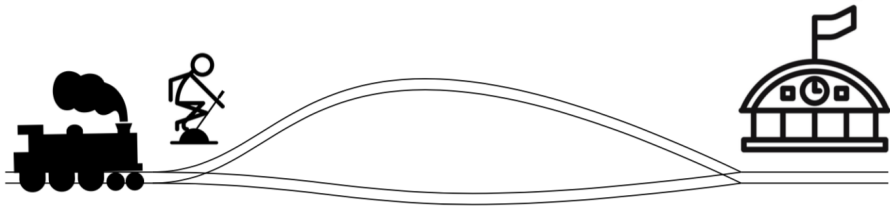- Hence $\neg C > \neg(\neg C \text{ cause } E)$

If C caused E, then, had C not occurred, the absence of C wouldn't have caused E.

$$C \text{ cause } E \quad \Rightarrow \quad \neg C > \neg(\neg C \text{ cause } E)$$

A nice result: Sartorio's principle is automatically satisfied when the effect counterfactually depends on the cause.

- Suppose $\neg C > \neg E$
- $\neg C$ cause E entails E,
- By contraposition, $\neg E$ entails $\neg(\neg C$ cause $E)$
- If X entails Y then $A > X$ entails $A > Y$
- Hence $\neg C > \neg(\neg C$ cause $E)$

Sartorio's Principle

If C caused E, then, had C not occurred, the absence of C wouldn't have caused E.

$$C \text{ cause } E \quad \Rightarrow \quad \neg C > \neg(\neg C \text{ cause } E)$$

A nice result: Sartorio's principle is automatically satisfied when the effect counterfactually depends on the cause.

- Suppose $\neg C > \neg E$
- $\neg C$ cause E entails E,
- By contraposition, $\neg E$ entails $\neg(\neg C$ cause $E)$
- If X entails Y then $A > X$ entails $A > Y$
- Hence $\neg C > \neg(\neg C$ cause $E)$

Counterfactual dependence is one way to make a difference.

But, as Suzy shows, it is not the only way to make a difference.

Plan

Suppose we have a semantics of *cause*, call it *proto-cause*, that does not account for the switches case.

Pulling the lever proto-caused the train to reach the station.

Suppose we have a semantics of *cause*, call it *proto-cause*, that does not account for the switches case.
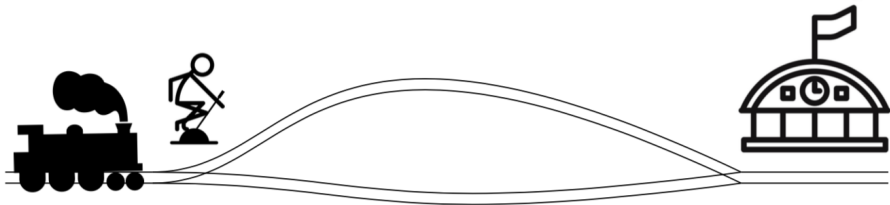
Pulling the lever proto-caused the train to reach the station.

Challenge

How do we amend *proto-cause* to predict that pulling the lever did **not** cause the train to reach the station?

Sartorio's Principle

If C caused E, then, had C not occurred, the absence of C wouldn't have caused E.

$$C \text{ cause } E \quad \Rightarrow \quad \neg C > \neg(\neg C \text{ cause } E)$$

If C caused E, then, had C not occurred, the absence of C wouldn't have caused E.

$$C \text{ cause } E \quad \Rightarrow \quad \neg C > \neg(\neg C \text{ cause } E)$$

$$x \geq 3x - 2$$
$$x + 2 \geq 3x$$
$$\frac{x + 2}{3} - x \geq 0$$

If C caused E, then, had C not occurred, the absence of C wouldn't have caused E.

$$C \text{ cause } E \quad \Rightarrow \quad \neg C > \neg(\neg C \text{ cause } E)$$

$$x \geq 3x - 2$$
$$x + 2 \geq 3x$$
$$\frac{x + 2}{3} - x \geq 0$$

Problem

- The operations of arithmetic have inverses (addition/subtraction; multiplication/division)
- Logical operations do not

Definition

Let $A[C/B]$ be the result of replacing every occurrence of B in A with C.

Example

$$((p \lor q) \land \neg q)[r/q] \quad = \quad (p \lor r) \land \neg r.$$

The Perfection Principle.

For any sentences C and E, there is a sentence X such that C *cause* E entails $C > X$ and $\neg(C > X)[\neg C/C]$.

The idea: X is the difference in 'making a difference'.

Sartorio's Principle

C cause E $\Rightarrow$ $\neg C > \neg(\neg C$ cause $E)$

The Perfection Principle

For all C and E, there is an X such that
C cause E $\Rightarrow$ $(C > X) \land \neg(C > X)[\neg C/C]$

Let $A \diamond\!\!\!\rightarrow C$ abbreviate $\neg(A > \neg C)$.

$A > C$ says: every A-world in the relevant domain is a C-world.
$A \diamond\!\!\!\rightarrow C$ says: some A-world in the relevant domain is a C-world.

(6)   a.   **Nonempty domains.** $A > C$ entails $A \diamond\!\!\!\rightarrow C$.

      b.   **Stability.** C *cause* E entails $C > (C$ *cause* $E)$.

      c.   **Idempotence.** $A \diamond\!\!\!\rightarrow C$ entails $A > (A \diamond\!\!\!\rightarrow C)$.

      d.   **Right weakening.**
         If C entails C' then $A > C$ entails $A > C'$.

      e.   If C *cause* E is true, then C is not a subsentence of E.

Let $A \diamondsuit\!\!\!\rightarrow C$ abbreviate $\neg(A > \neg C)$.

$A > C$ says: every A-world in the relevant domain is a C-world.
$A \diamondsuit\!\!\!\rightarrow C$ says: some A-world in the relevant domain is a C-world.

(6)   a.   **Nonempty domains.** $A > C$ entails $A \diamondsuit\!\!\!\rightarrow C$.

        b.   **Stability.** C *cause* E entails $C > (C$ *cause* $E)$.

        c.   **Idempotence.** $A \diamondsuit\!\!\!\rightarrow C$ entails $A > (A \diamondsuit\!\!\!\rightarrow C)$.

        d.   **Right weakening.**
            If C entails C' then $A > C$ entails $A > C'$.

        e.   If C *cause* E is true, then C is not a subsentence of E.

Theorem

Given the assumptions in (6), Sartorio's Principle is equivalent to the Perfection Principle.

Sartorio's Principle

$C \text{ cause } E \;\Rightarrow\; \neg C > \neg(\neg C \text{ cause } E)$

The Perfection Principle

For all C and E, there is an X such that
$C \text{ cause } E \;\Rightarrow\; (C > X) \;\wedge\; \neg(C > X)[\neg C/C]$

Suppose Sartorio's Principle. Pick any sentences C and E and take $X = (C\ cause\ E)$. Then by Stability, $C\ cause\ E$ entails $C > X$. We also have the following chain of implications.

$$
\begin{aligned}
C\ cause\ E \quad &\Rightarrow \quad \neg C > \neg(\neg C\ cause\ E) && \text{(Sartorio's Principle)} \\
&\Rightarrow \quad \neg(\neg C > (\neg C\ cause\ E)) && \text{(Nonempty domains)} \\
&\Rightarrow \quad \neg(C > (C\ cause\ E))[\neg C/C] && \\
&&& \hspace{-3cm}(C \text{ is not a subsentence of } E) \\
&\Rightarrow \quad \neg(C > X)[\neg C/C] && (X = C\ cause\ E)
\end{aligned}
$$

Hence $C\ cause\ E$ entails $C > X$ and $\neg(C > X)[\neg C/C]$.

Suppose the Perfection Principle. So $\neg C$ *cause* E entails $(C > X)[\neg C/C]$. Then by contraposition we have (†): $\neg(C > X)[\neg C/C]$ entails $\neg(\neg C$ *cause* $E)$. Observe the following chain of implications.

$$
\begin{array}{lll}
C \text{ cause } E & \Rightarrow & \neg(C > X)[\neg C/C] \qquad \text{(Perfection Principle)} \\
& \Rightarrow & \neg(\neg C > X[\neg C/C]) \qquad \text{(Definition of } [\neg C/C]) \\
& \Rightarrow & \neg C \diamondsuit\!\!\rightarrow \neg X[\neg C/C] \qquad \text{(Definition of } \diamondsuit\!\!\rightarrow) \\
& \Rightarrow & \neg C > (\neg C \diamondsuit\!\!\rightarrow \neg X[\neg C/C]) \qquad \text{(Idempotence)} \\
& \Rightarrow & \neg C > \neg(\neg C > X[\neg C/C]) \qquad \text{(Definition of } \diamondsuit\!\!\rightarrow) \\
& \Rightarrow & \neg C > \neg(C > X)[\neg C/C] \qquad \text{(Definition of } [\neg C/C]) \\
& \Rightarrow & \neg C > \neg(\neg C \text{ cause } E) \qquad \text{(† and right weakening)}
\end{array}
$$

- Presumably, causal claims say something about what would happen if the cause occurred, or say something about what would happen if the cause did not occur.

  For some X, C *proto-cause* E entails $C > X$ or $\neg(C > X)[\neg C/C]$.

- Presumably, causal claims say something about what would happen if the cause occurred, or say something about what would happen if the cause did not occur.

  For some X, C *proto-cause* E entails $C > X$ or $\neg(C > X)[\neg C/C]$.

- Presumably, causal claims say something about what would happen if the cause occurred, or say something about what would happen if the cause did not occur.

  For some X, C *proto-cause* E entails $C > X$ or $\neg(C > X)[\neg C/C]$.

- Our theorem gives us a way to turn *proto-cause* into *cause*.

- Presumably, causal claims say something about what would happen if the cause occurred, or say something about what would happen if the cause did not occur.

  For some X, C *proto-cause* E entails $C > X$ or $\neg(C > X)[\neg C / C]$.

- Our theorem gives us a way to turn *proto-cause* into *cause*.

- Presumably, causal claims say something about what would happen if the cause occurred, or say something about what would happen if the cause did not occur.

  For some X, C *proto-cause* E entails $C > X$ or $\neg(C > X)[\neg C/C]$.

- Our theorem gives us a way to turn *proto-cause* into *cause*.
- **Case 1**. If C *proto-cause* E entails $C > X$ then add $\neg(C > X)[\neg C/C]$:

  C *cause* E if and only if C *proto-cause* $E \wedge \neg(\neg C > X)[\neg C/C]$

- Presumably, causal claims say something about what would happen if the cause occurred, or say something about what would happen if the cause did not occur.

  For some X, C *proto-cause* E entails $C > X$ or $\neg(C > X)[\neg C/C]$.

- Our theorem gives us a way to turn *proto-cause* into *cause*.
- **Case 1**. If C *proto-cause* E entails $C > X$ then add $\neg(C > X)[\neg C/C]$:

  C *cause* E if and only if C *proto-cause* $E \wedge \neg(\neg C > X)[\neg C/C]$

- **Case 2**. If C *proto-cause* E entails $\neg(C > X)[\neg C/C]$ then add $C > X$.

  C *cause* E if and only if C *proto-cause* $E \wedge C > X$

Plan

- Lewis (1973, p. 536): event *e* causally depends on an event *c* just in case the following two counterfactuals are true: if *c* had occurred, *e* would have occurred.

- Wright (1985, 2011) proposes the NESS (Necessary Element of a Sufficient Set) test for causation:

  C is a cause of *C* just in case there is a set of conditions *B* that is jointly sufficient for *E*, but $B - \{C\}$ is not sufficient for *E*.

- Mackie's INUS condition (Mackie 1974):

  A cause is "an insufficient but non-redundant part of a condition which is itself unnecessary but sufficient for the result" (Mackie 1974, p. 64).

  This implies that is a background B such that B is sufficient for E but $B - \{C\}$ is not sufficient for E.

- Beckers (2016) use a notion of *production*:

  C is an cause of E just in case (after intervening to make C true), C produces E, and after intervening to make $\neg C$ true, $\neg C$ does not produce E.

Plan

Certainly, it seems to be the case that an inference can, historically, become part of semantic representation in the strict sense; thus, the development of the English conjunction since from a purely temporal word to a marker of causation can be interpreted as a change from a principle of invited inference associated with since (by virtue of its temporal meaning) to a piece of the semantic content of since.

(Geis and Zwicky 1971, pp. 565–566)

Iatridou (1993, 2021) observes that *then* in conditionals takes on a further meaning. She offers the following examples, which are unacceptable with *then* but fine without it.

(7) a. If I may be frank (*then) you are not looking good today.
     b. If John is dead or alive (*then) Bill will find him.
     c. If he were the last man on earth (*then) she wouldn't marry him.
     d. Even if you give me a million dollars (*then) I will not sell you my piano.

Iatridou (1993, 2021) observes that *then* in conditionals takes on a further meaning. She offers the following examples, which are unacceptable with *then* but fine without it.

(7)   a.   If I may be frank (*then) you are not looking good today.

        b.   If John is dead or alive (*then) Bill will find him.

        c.   If he were the last man on earth (*then) she wouldn't marry him.

        d.   Even if you give me a million dollars (*then) I will not sell you my piano.

Where $A > C$ denotes the conditional construction, Iatridou (1993) proposes that *if A then C* asserts $A > C$ and presupposes $\neg(\neg A > C)$.

Iatridou (1993, 2021) observes that *then* in conditionals takes on a further meaning. She offers the following examples, which are unacceptable with *then* but fine without it.

(7)   a.   If I may be frank (*then) you are not looking good today.
      b.   If John is dead or alive (*then) Bill will find him.
      c.   If he were the last man on earth (*then) she wouldn't marry him.
      d.   Even if you give me a million dollars (*then) I will not sell you my piano.

Where $A > C$ denotes the conditional construction, Iatridou (1993) proposes that *if A then C* asserts $A > C$ and presupposes $\neg(\neg A > C)$.

The Perfection Principle.

For any sentences C and E, there is a sentence X such that C *cause* E entails $C > X$ and $\neg(C > X)[\neg C/C]$.

Iatridou (1993, 2021) observes that *then* in conditionals takes on a further meaning. She offers the following examples, which are unacceptable with *then* but fine without it.

(7) a. If I may be frank (*then) you are not looking good today.
     b. If John is dead or alive (*then) Bill will find him.
     c. If he were the last man on earth (*then) she wouldn't marry him.
     d. Even if you give me a million dollars (*then) I will not sell you my piano.

Where $A > C$ denotes the conditional construction, Iatridou (1993) proposes that *if A then C* asserts $A > C$ and presupposes $\neg(\neg A > C)$.

The Perfection Principle.

For any sentences C and E, there is a sentence X such that C *cause* E entails $C > X$ and $\neg(C > X)[\neg C/C]$.

The similarity is striking!

CAUSATION
AND MODALITY

McHugh (2023): the difference in 'making a difference' is that the cause produced the effect:

$$X = (C \ produce \ E)$$

Dean McHugh

$$C \ \wedge \ C \gg (C \ produce \ E) \ \wedge \ \neg(\neg C \gg (\neg C \ produce \ E))$$

- Pulling the lever produced the train to reach the station, but if the engineer hadn't pulled the lever, not pulling the lever would have also produced the train to reach the station.

- Pulling the lever produced the train to reach the station, but if the engineer hadn't pulled the lever, not pulling the lever would have also produced the train to reach the station.

- Suzy throwing her rock produced the window to break, and if she hadn't thrown her rock, not throwing the rock would not have produced the window to break.

Summary

- Sartorio's Principle offers a principled way to distinguish the Billy and Suzy case from switching cases.

Summary

- Sartorio's Principle offers a principled way to distinguish the Billy and Suzy case from switching cases.
- But given the Principle's logical structure, one cannot simply add it to existing semantics of *cause*.

Summary

- Sartorio's Principle offers a principled way to distinguish the Billy and Suzy case from switching cases.
- But given the Principle's logical structure, one cannot simply add it to existing semantics of *cause*.
- Our theorem provides an automatic way to add Sartorio's Principle to semantic theories of *cause*.

Thank you!

References I

📄 Anscombe, Gertrude Elizabeth Margaret (1971). *Causality and determination: An inaugural lecture*. CUP Archive.

📄 Beckers, Sander (2016). Actual Causation: Definitions and Principles. PhD thesis. KU Leuven. URL: `https://limo.libis.be/primo-explore/fulldisplay?docid=LIRIAS1656621&context=L&vid=Lirias&search_scope=Lirias&tab=default_tab&lang=en_US`.

📄 Geis, Michael L. and Arnold M. Zwicky (1971). On invited inferences. *Linguistic inquiry* 2.4, pp. 561–566. URL: `www.jstor.org/stable/4177664`.

📄 Hall, Ned (2000). Causation and the Price of Transitivity. *Journal of Philosophy* 97.4, pp. 198–222. DOI: 10.2307/2678390.

📄 — (2004). "Two concepts of causation". *Causation and counterfactuals*. Ed. by John Collins, Ned Hall, and Paul Laurie. MIT Press, pp. 225–276.

📄 Hall, Ned and Laurie A. Paul (2003). Causation and preemption. *Philosophy of Science Today*, pp. 100–130.

References II

📄 Halpern, Joseph Y (2016). *Actual Causality*. MIT Press.

📄 Halpern, Joseph Y and Judea Pearl (2005). Causes and explanations: A structural-model approach. Part I: Causes. *The British journal for the philosophy of science* 56.4, pp. 843–887. DOI: 10.1093/bjps/axi147.

📄 Hitchcock, Christopher (2012). Events and times: A case study in means-ends metaphysics. *Philosophical Studies* 160.1, pp. 79–96. DOI: 10.1007/s11098-012-9909-4.

📄 Horn, Laurence R (1969). A presuppositional analysis of *only* and *even*. *Proceedings from the Annual Meeting of the Chicago Linguistic Society*. Vol. 5. Chicago Linguistic Society, pp. 98–107.

📄 Iatridou, Sabine (1993). On the contribution of conditional then. *Natural language semantics* 2.3, pp. 171–199.

📄 — (2021). "Grammar matters". *Conditionals, Paradox, and Probability: Themes from the Philosophy of Dorothy Edgington*. Ed. by Lee Walters and John Hawthorne. Oxford University Press. DOI: 10.1093/oso/9780198712732.003.0008.

Lewis, David (1973). Causation. *Journal of Philosophy* 70.17, pp. 556–567. DOI: 10.2307/2025310.

— (2000). Causation as Influence. *Journal of Philosophy* 97.4, pp. 182–197. DOI: 10.2307/2678389.

Mackie, John L (1974). *The cement of the universe: A study of causation*. Clarendon Press. DOI: 10.1093/0198246420.001.0001.

McHugh, Dean (2023). Causation and Modality: Models and Meanings. PhD thesis. University of Amsterdam. URL: https://eprints.illc.uva.nl/id/eprint/2243.

Paul, L. A. (1998). Keeping Track of the Time: Emending the Counterfactual Analysis of Causation. *Analysis* 58.3, pp. 191–198. DOI: 10.1111/1467-8284.00121.

📄 Rooth, Mats (1985). Association with focus. PhD thesis. University of Massachusetts, Amherst. URL: url=https%5C%3A%5C%2F%5C%2Fecommons.cornell.edu%5C%2Fbitstream%5C%2Fhandle%5C%2F1813%5C%2F28568%5C%2FRooth-1985-PhD.pdf&usg=AOvVaw2X_mAUuPyshpTVYY-Q3mK_.

📄 Sartorio, Carolina (2005). Causes As Difference-Makers. *Philosophical Studies* 123.1, pp. 71–96. DOI: 10.1007/s11098-004-5217-y.

📄 Wright, Richard (1985). Causation in tort law. *California Law Review* 73.6, pp. 1735–1828. DOI: 10.2307/3480373.

📄 — (2011). "The NESS account of natural causation: a response to criticisms". *Perspectives on Causation*. Ed. by Richard Goldberg. Hart Publishing, pp. 13–66.

(8) *C cause E* and *E because C* entail …
   a. Existential difference-making: $\neg(\neg C > (\neg C \text{ produce } E))$
   b. Universal difference-making: $\neg C > \neg(\neg C \text{ produce } E)$

(9)   a.   Reyna was born at Royal Bolton Hospital but received a Danish passport because her mother was born in Copenhagen.

           b.   He has an American passport because he was born in Boston.

           c.   I think I was laid off because I'm 56 years old.

           d.   Naama Issachar … could spend up to seven-and-a-half years in a Russian prison because 9.5 grams of cannabis were found in her possession during a routine security check.

           e.   A 90-day study in 8 adults found that supplementing a standard diet with 1.3 cups (100 grams) of fresh coconut daily caused significant weight loss.

Local Man Paralysed After Eating 413 Chicken Nuggets

swissguy25 🇨🇭 3h
So the Limit is 412

(10)   If he hadn't eaten 413 chicken nuggets, he wouldn't have been paralysed.
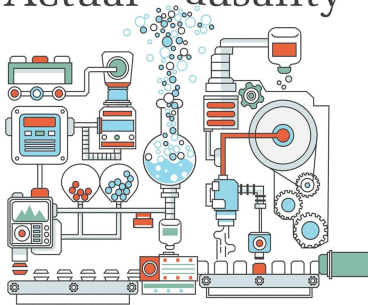
Actual Causality
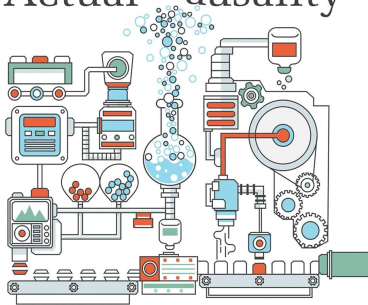
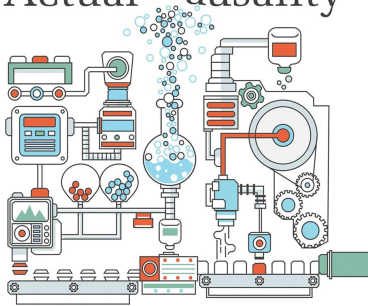Joseph Y. Halpern

Halpern (2016), *Actual Causality*:

C is an actual cause of E just in case

1. C and E actually occurred.

Actual Causality

Joseph Y. Halpern

Halpern (2016), *Actual Causality*:

C is an actual cause of *E* just in case

1. *C* and *E* actually occurred.
2. There is a set of variables such that, holding them fixed at their actual values, if the cause had not occurred, the effect would not have occurred.
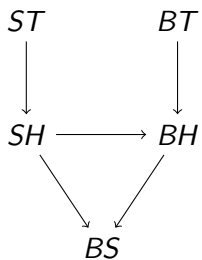
Actual Causality

Joseph Y. Halpern

Halpern (2016), *Actual Causality*:

C is an actual cause of E just in case

1. C and E actually occurred.
2. There is a set of variables such that, holding them fixed at their actual values, if the cause had not occurred, the effect would not have occurred.
3. C is minimal: no proper subset of C satisfies (1) and (2).

Figure: Halpern's model of the Billy and Suzy case (2016, p. 31)

Halpern's account of the Billy and Suzy case

$$ST = 1 \qquad BT = 1$$

$$SH = 1 \longrightarrow BH = 0$$

$$BS = 1$$

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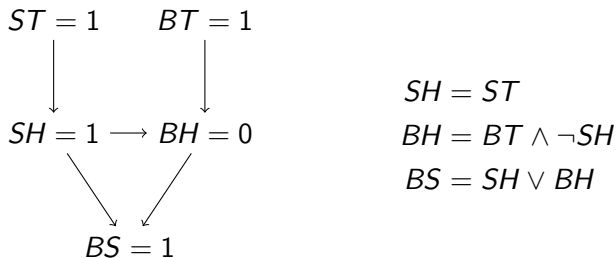$$BS = SH \vee BH$$

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

Halpern's account of the Billy and Suzy case
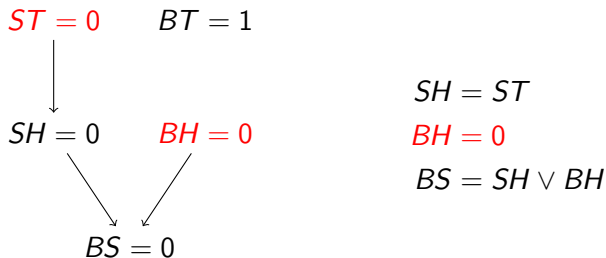
Figure: Halpern's model of *Late preemption* (2016, p. 31)

Two models of the switching scenario
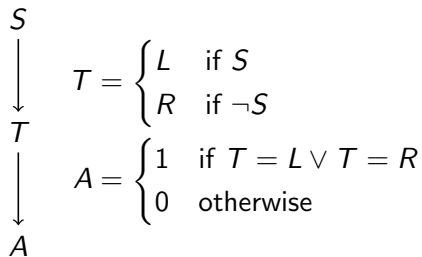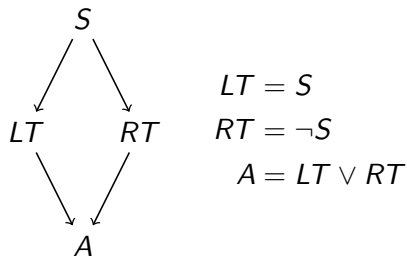
S

$T = \begin{cases} L & \text{if } S \\ R & \text{if } \neg S \end{cases}$

$A = \begin{cases} 1 & \text{if } T = L \vee T = R \\ 0 & \text{otherwise} \end{cases}$

T

A

(a) One-variable model

S

$LT \qquad RT$

A

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

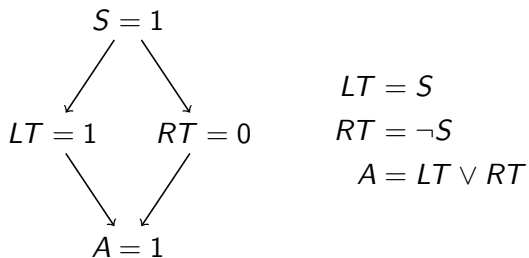(b) Two-variable model

The two-variable model

Figure: Two-variable model
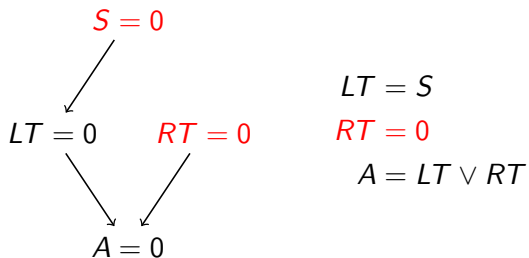
The two-variable model

Figure: Two-variable model

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.
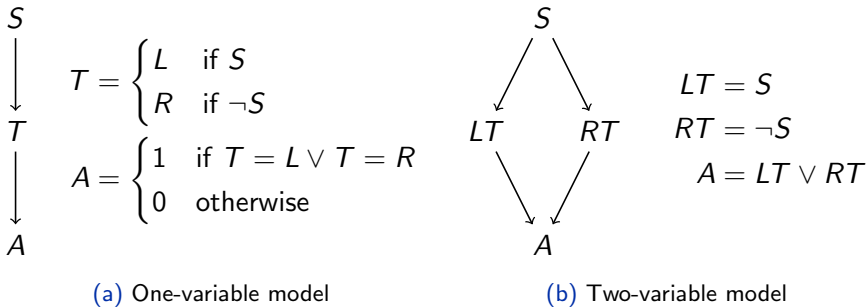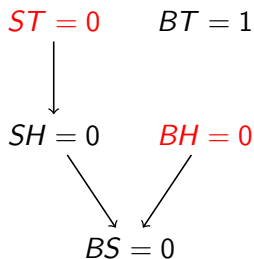
$$S$$
$$\downarrow$$
$$T$$
$$\downarrow$$
$$A$$

$$T = \begin{cases} L & \text{if } S \\ R & \text{if } \neg S \end{cases}$$

$$A = \begin{cases} 1 & \text{if } T = L \vee T = R \\ 0 & \text{otherwise} \end{cases}$$

(a) One-variable model

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

(b) Two-variable model

Figure: Two models of the switching scenario

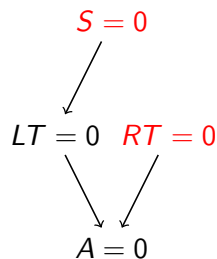In the two-variable model, one can intervene to make

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

That is, interventions can make the train disappear from the tracks!
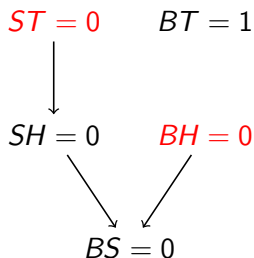
The two-variable model

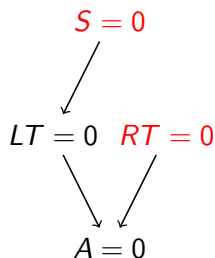(a) Witness to Suzy causing the window to break

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

The two-variable model

(a) Witness to Suzy causing the window to break

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

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

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.*

Halpern's solution to the Billy and Suzy case is too sensitive to the choice of model.

Sartorio's Principle offers a more robust solution.

E depends causally on C iff C occurs, E occurs, and if C had not occurred, then E would not have occurred at all, or would have occurred later than the time that it actually did occur.

(Paul 1998, p. 193)

Suppose it were alleged that since we are all mortal, there is no such thing as a cause of death. Without the hanging that allegedly caused the death of Ned Kelly, for instance, he would sooner or later have died anyway. Yes. But he would have died a different death, and the event that actually was Kelly's death would never have occurred.

(Lewis 2000, pp. 185)

This proposal does not abandon the strategy of fragility, but corrects it. Instead of supposing that the event itself is fragile—which would fly in the face of much of our ordinary talk—we instead take a tailor-made fragile proposition about that event and its time.

If we stopped here, we would be building into our analysis an asymmetry between hasteners and delayers. ... To restore symmetry between hastening and delaying, we need only replace the words 'or would have occurred later than the time that it actually did occur' by the words 'or would have occurred at a time different from the time that it actually did occur'. I favor this further emendation. (As does Paul.) But I think we should go further still. What is so special about time?

(Lewis 2000, p. 187)

We could further emend our analysis to require dependence of how and when and whether upon whether: without C, E would not have occurred at all, or would have occurred at a time different from the time that it actually did occur, or would have occurred in a manner different from the manner in which it actually did occur.

(Lewis 2000, p. 187)

Lewis on fragility

We could further emend our analysis to require dependence of how and when and whether upon whether: without C, E would not have occurred at all, or would have occurred at a time different from the time that it actually did occur, or would have occurred in a manner different from the manner in which it actually did occur.
(Lewis 2000, p. 187)

- Did Suzy's throwing her rock change the manner in which the bottle broke?
- Did the engineer pulling the lever change the manner in which the train reached the station?

(11)   The bottle broke because Suzy threw her rock at it.

- Suzy throwing her rock at the bottle is sufficient for it to break,
- but not sufficient for it to break in the way that it did.

To keep a uniform notion of sufficiency, if we apply event fragility to the case where the difference-making condition, we should also apply it to the sufficiency condition.

We would like clear predictions for clear judgements.
Are

(12)   a.   Suzy throwing her rock caused the bottle to break.
        b.   The engineer pulling the lever did not cause the train to
              reach the station.

clearly true? If so, we would like this to be a clear prediction of our
account.

if a meeting is originally scheduled for Monday at noon, and then re-scheduled for Tuesday at noon, is the meeting that occurs on Tuesday at noon the very same meeting that would have occurred on Monday? That is, was the meeting postponed, strictly speaking, or was the original meeting cancelled and a different meeting scheduled for Tuesday?

(Hitchcock 2012, p. 83)

Suppose the Athenian citizens vote to put Socrates to death, but leave it to the executioner to decide when he has to die. The executioner was planning a year-long trip to Babylon, but his boat was destroyed in a storm. Socrates died in 399 BCE, but if the executioner's boat hadn't been destroyed Socrates would have died a year later, in 398 BCE. Consider:

(13)   a.   Socrates died because the executioner's boat was destroyed.
         b.   The fact that the executioner's boat was destroyed caused Socrates to die.

(14)   a.   Socrates died in 399 BCE because the executioner's boat was destroyed.
         b.   The fact that the executioner's boat was destroyed caused Socrates to die in 399 BCE.

Imagine that the executioner had only one dose of hemlock left, designated for another prisoner. The Athenians originally wished to throw Socrates off a cliff. However, the other prisoner was released, so the hemlock was given to Socrates instead. Consider:

(15)   a.   Socrates died because the other prisoner was released.
          b.   The other prisoner's release caused Socrates to die.

noncauses can easily make a difference to the time and manner of an event's occurrence—a gust of wind that alters the course of Suzy's rock ever so slightly, for example

(Hall 2004, p. 237)

(16) a. Reyna received a Danish passport because her mother was born in Copenhagen.
      b. The bottle broke because Suzy threw her rock at it.
      c. Socrates died because he drank poison.

(16) a.   Reyna received a Danish passport because her mother was
           born in Copenhagen.
      b.   The bottle broke because Suzy threw her rock at it.
      c.   Socrates died because he drank poison.

(17) a.   Reyna only received a Danish passport because her mother
           was born in Copenhagen.
      b.   The bottle only broke because Suzy threw her rock at it.
      c.   Socrates only died because he drank poison.

(18)   a.   The reason Reyna received a Danish passport is that her mother was born in Copenhagen.

           b.   The reason the bottle broke is that Suzy threw her rock at it.

           c.   The reason Socrates died is that he drank poison.

(18) a. The reason Reyna received a Danish passport is that her
          mother was born in Copenhagen.
      b. The reason the bottle broke is that Suzy threw her rock at it.
      c. The reason Socrates died is that he drank poison.

(19) a. The only reason Reyna received a Danish passport is that her
          mother was born in Copenhagen.
      b. The only reason the bottle broke is that Suzy threw her rock
          at it.
      c. The only reason Socrates died was that he drank poison.

Only is interpreted with respect to a set of alternatives (Rooth 1985).

(20)   **Meaning of only** (Horn 1969). For any sentence S and set of
        sentences *Alt*, $only_{Alt}$ S asserts that for every $A \in Alt$, if S does
        not entail A then A is false.

Only is interpreted with respect to a set of alternatives (Rooth 1985).

(20) **Meaning of only** (Horn 1969). For any sentence S and set of sentences *Alt*, *only$_{Alt}$ S* asserts that for every $A \in Alt$, if S does not entail A then A is false.

(21) a. I only introduced BILL to Sue.

     b. I only introduced Bill to SUE.

Only is interpreted with respect to a set of alternatives (Rooth 1985).

(20)   **Meaning of only** (Horn 1969). For any sentence S and set of sentences *Alt*, *only$_{Alt}$ S* asserts that for every $A \in Alt$, if S does not entail A then A is false.

(21)   a.   I only introduced BILL to Sue.
        b.   I only introduced Bill to SUE.

In (21a), *only* negates alternatives of the form *I introduced x to Sue*, saying I didn't introduce anyone but Bill to Sue, while in (21b) it negates alternatives of the form *I introduced Bill to x*, saying that I didn't introduce Bill to anyone but Sue.

Suppose both of Reyna's parents were born in Copenhagen, but in Reyna's case the law only allows her mother, not her father, to pass on citizenship to her. In that case

(22)   Reyna only received a Danish passport because her mother was born in Copenhagen.

may have a single alternative:

(23)   Reyna received a Danish passport because her father was born in Copenhagen.

Proposal: the set of alternatives can also be all other *because*-clauses.

$$ALT(E \text{ because } C) = \{E \text{ because } D : D \text{ is a sentence}\}$$

(24)   Reyna only received a Danish passport because her mother was born in Copenhagen.

(25)   Reyna received a Danish passport because her mother was born in Denmark.

- (26) does not entail (27).
- In a world where only those born in Copenhagen receive Danish passports, (26) is true but (27) is false.
- (In that world (27) fails the sufficiency requirement.)

Given this set of alternatives,

(26)   Reyna only received a Danish passport because her mother was
         born in Copenhagen.

asserts that

(27)   Reyna received a Danish passport because her mother was born in
         Denmark.

is false.

Given this set of alternatives,

(26)   Reyna only received a Danish passport because her mother was born in Copenhagen.

asserts that

(27)   Reyna received a Danish passport because her mother was born in Denmark.

is false.

$$\neg(E \text{ because } D)$$
$$\Leftrightarrow \quad \neg\Big(D \wedge \big(D \gg (D \text{ produce } E)\big) \wedge \neg\big(\neg D \gg (\neg D \text{ produce } E)\big)\Big)$$
$$\Leftrightarrow \quad \neg D \vee \neg\big(D \gg (D \text{ produce } E)\big) \vee \big(\neg D \gg (\neg D \text{ produce } E)\big)$$

The first and third disjuncts are false.
The second disjunct is also false: Reyna's mother being born in Copenhagen is indeed sufficient for that to produce Reyna to receive a Danish passport.

(28)   The bottle broke only because Suzy threw her rock at it.

(29)   ⇒ ¬(The bottle broke because Suzy or Billy threw a rock at it.)

  ¬(*Suzy or Billy throw*)
∨ ¬((*Suzy or Billy throw*) ≫ ((*Suzy or Billy throw*) *produce bottle break*))
∨ (¬(*Suzy or Billy throw*) ≫ (¬(*Suzy or Billy throw*) *produce bottle break*))

But Suzy or Billy throwing is sufficient for the bottle to break.

A problem for the event fragility strategy

Suzy or Billy throwing is sufficient for the bottle to break, but not sufficient for it to break in the way that it did.

If we adopt event fragility, we lose a natural and straightforward account of why (28) is unacceptable.

References I

📄 Anscombe, Gertrude Elizabeth Margaret (1971). *Causality and determination: An inaugural lecture*. CUP Archive.

📄 Beckers, Sander (2016). Actual Causation: Definitions and Principles. PhD thesis. KU Leuven. URL: https://limo.libis.be/primo-explore/fulldisplay?docid=LIRIAS1656621&context=L&vid=Lirias&search_scope=Lirias&tab=default_tab&lang=en_US.

📄 Geis, Michael L. and Arnold M. Zwicky (1971). On invited inferences. *Linguistic inquiry* 2.4, pp. 561–566. URL: www.jstor.org/stable/4177664.

📄 Hall, Ned (2000). Causation and the Price of Transitivity. *Journal of Philosophy* 97.4, pp. 198–222. DOI: 10.2307/2678390.

📄 — (2004). "Two concepts of causation". *Causation and counterfactuals*. Ed. by John Collins, Ned Hall, and Paul Laurie. MIT Press, pp. 225–276.

📄 Hall, Ned and Laurie A. Paul (2003). Causation and preemption. *Philosophy of Science Today*, pp. 100–130.

References II

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

Halpern, Joseph Y and Judea Pearl (2005). Causes and explanations: A structural-model approach. Part I: Causes. *The British journal for the philosophy of science* 56.4, pp. 843–887. DOI: 10.1093/bjps/axi147.

Hitchcock, Christopher (2012). Events and times: A case study in means-ends metaphysics. *Philosophical Studies* 160.1, pp. 79–96. DOI: 10.1007/s11098-012-9909-4.

Horn, Laurence R (1969). A presuppositional analysis of *only* and *even*. *Proceedings from the Annual Meeting of the Chicago Linguistic Society*. Vol. 5. Chicago Linguistic Society, pp. 98–107.

Iatridou, Sabine (1993). On the contribution of conditional then. *Natural language semantics* 2.3, pp. 171–199.

— (2021). "Grammar matters". *Conditionals, Paradox, and Probability: Themes from the Philosophy of Dorothy Edgington*. Ed. by Lee Walters and John Hawthorne. Oxford University Press. DOI: 10.1093/oso/9780198712732.003.0008.

Lewis, David (1973). Causation. *Journal of Philosophy* 70.17, pp. 556–567. DOI: 10.2307/2025310.

— (2000). Causation as Influence. *Journal of Philosophy* 97.4, pp. 182–197. DOI: 10.2307/2678389.

Mackie, John L (1974). *The cement of the universe: A study of causation*. Clarendon Press. DOI: 10.1093/0198246420.001.0001.

McHugh, Dean (2023). Causation and Modality: Models and Meanings. PhD thesis. University of Amsterdam. URL: https://eprints.illc.uva.nl/id/eprint/2243.

Paul, L. A. (1998). Keeping Track of the Time: Emending the Counterfactual Analysis of Causation. *Analysis* 58.3, pp. 191–198. DOI: 10.1111/1467-8284.00121.

References IV

Rooth, Mats (1985). Association with focus. PhD thesis. University of Massachusetts, Amherst. URL: url=https%5C%3A%5C%2F%5C%2Fecommons.cornell.edu%5C%2Fbitstream%5C%2Fhandle%5C%2F1813%5C%2F28568%5C%2FRooth-1985-PhD.pdf&usg=AOvVaw2X_mAUuPyshpTVYY-Q3mK_.

Sartorio, Carolina (2005). Causes As Difference-Makers. *Philosophical Studies* 123.1, pp. 71–96. DOI: 10.1007/s11098-004-5217-y.

Wright, Richard (1985). Causation in tort law. *California Law Review* 73.6, pp. 1735–1828. DOI: 10.2307/3480373.

— (2011). "The NESS account of natural causation: a response to criticisms". *Perspectives on Causation*. Ed. by Richard Goldberg. Hart Publishing, pp. 13–66.