

Responses to the problem of overdetermination

Philosophical foundations of explanation

Dean McHugh

Institute of Logic, Language and Computation
University of Amsterdam

January 15, 2024



INSTITUTE FOR LOGIC,
LANGUAGE AND COMPUTATION



UNIVERSITY
OF AMSTERDAM

Plan

- 1 Structural causal models
- 2 Comparison with Halpern's *Actual Causality*
- 3 Overdetermination via fragility
 - Cases where the effect would still have happened at the same time
- 4 Only because

Plan

- 1 Structural causal models
- 2 Comparison with Halpern's *Actual Causality*
- 3 Overdetermination via fragility
 - Cases where the effect would still have happened at the same time
- 4 Only because

Structural causal models

Definition (Structural causal model)

A structural causal model is a triple $M = (V, E, F)$ where

- V is a set of variables
- (V, E) is a directed acyclic graph
- F is a set of functions of the form

$$F_X : \mathcal{R}(pa_X) \rightarrow \mathcal{R}(X),$$

one for each endogenous (i.e. with a parent) variable $X \in V$.

Structural causal models

Definition (Structural causal model)

A structural causal model is a triple $M = (V, E, F)$ where

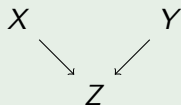
- V is a set of variables
- (V, E) is a directed acyclic graph
- F is a set of functions of the form

$$F_X : \mathcal{R}(pa_X) \rightarrow \mathcal{R}(X),$$

one for each endogenous (i.e. with a parent) variable $X \in V$.

Example

$$Z = X \vee Y$$



$$V = \{X, Y, Z\}$$

$$\mathcal{R}(U) = \{0, 1\} \text{ for all } U \in V$$

$$F_X(Y, Z) = 1 \text{ iff } Y = 1 \vee Z = 1$$

Figure: Structural causal model of an OR-gate.

Structural causal models

The value of an endogenous variable X is determined by the values of its parents, according to F_X

- Since F_X are **functions**, the dependence is **deterministic**
- Where $U = u$ is an assignment of values to the exogenous variables in V , we call u a *setting* or *context* for M
 - i.e. the values of the exogenous variables determine the values of all the variables

Interventions in structural causal models

Let $M = (V, E, F)$ be a structural causal model

Definition (Interventions as model surgery)

$M_{X=x}$ is the model $(V, E, F_{X=x})$ which results from replacing the equation for X in M with $X = x$ (that is, $F_{X=x} := (F \setminus \{F_X\}) \cup \{F'_X\}$ where $F'_X(y_1, y_2, \dots) = x$ for any values y_1, y_2, \dots of X 's parents).

Interventions in structural causal models

Let $M = (V, E, F)$ be a structural causal model

Definition (Interventions as model surgery)

$M_{X=x}$ is the model $(V, E, F_{X=x})$ which results from replacing the equation for X in M with $X = x$ (that is, $F_{X=x} := (F \setminus \{F_X\}) \cup \{F'_X\}$ where $F'_X(y_1, y_2, \dots) = x$ for any values y_1, y_2, \dots of X 's parents).

Definition (Truth conditions for interventions)

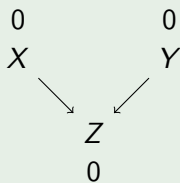
Let M be a structural causal model and u a setting of the exogenous variables.

$$M, u \models [X \leftarrow x]Y = y \quad \text{iff} \quad M_{X=x}, u \models Y = y$$

Example of an intervention

Example

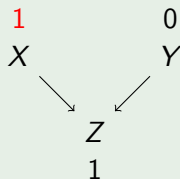
$$Z = X \vee Y$$



Example of an intervention

Example (Intervene to set $X = 1$)

$$Z = X \vee Y$$



Example of an intervention: a chain

Example

X 0



Y 0



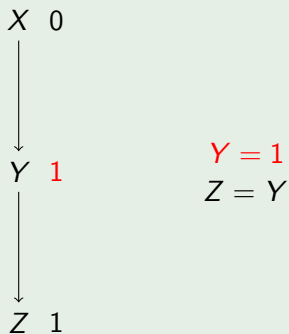
Z 0

$$Y = X$$

$$Z = Y$$

Intervene to set $Y = 1$

Example



Let M be the model above and $u = (0, 0, 0)$.

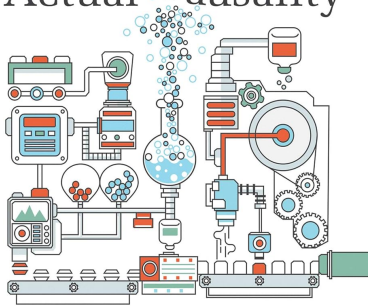
$$M, u \models [Y = 1]X = 0$$

$$M, u \models [Y = 1]Z = 1$$

Plan

- 1 Structural causal models
- 2 Comparison with Halpern's *Actual Causality*
- 3 Overdetermination via fragility
 - Cases where the effect would still have happened at the same time
- 4 Only because

Actual Causality



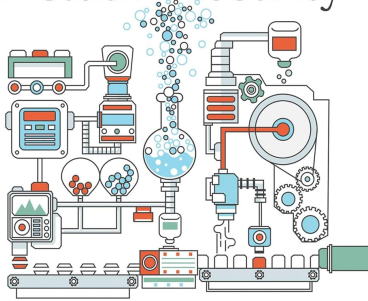
Joseph Y. Halpern

Halpern (2016), *Actual Causality*:

C is an actual cause of E just in case

- 1 C and E actually occurred.

Actual Causality



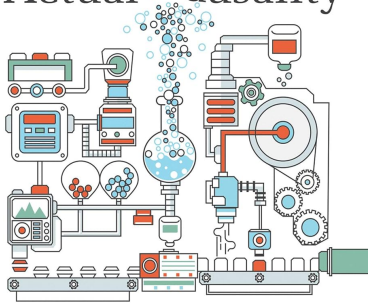
Joseph Y. Halpern

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

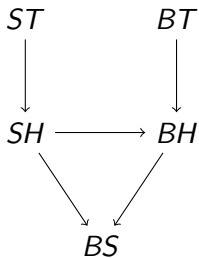


Joseph Y. Halpern

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).



$$SH = ST$$

$$BH = BT \wedge \neg SH$$

$$BS = SH \vee BH$$

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

Halpern's account of the Billy and Suzy case

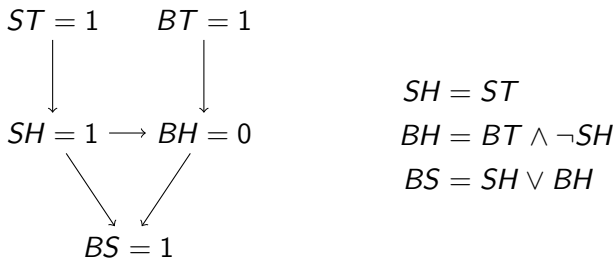


Figure: Halpern's model of *Late preemption* (2016, p. 31)

Halpern's account of the Billy and Suzy case

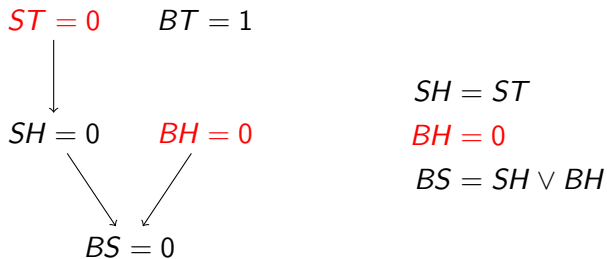
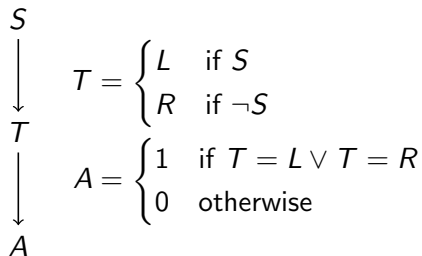
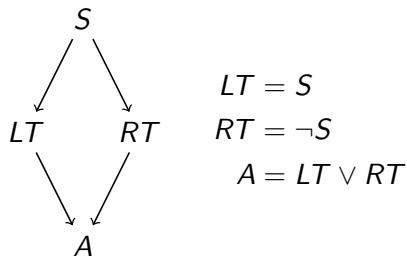


Figure: Halpern's model of *Late preemption* (2016, p. 31)

Two models of the switching scenario



(a) One-variable model



(b) Two-variable model

The two-variable model

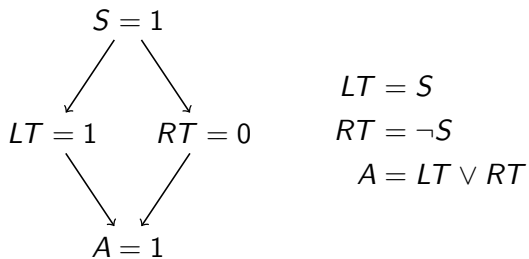


Figure: Two-variable model

The two-variable model

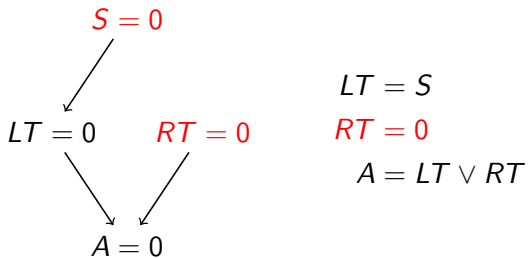
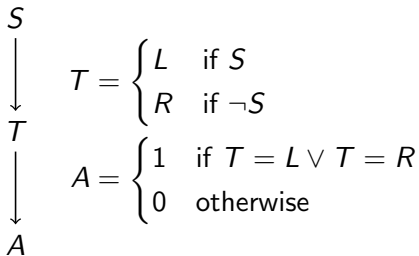


Figure: Two-variable model

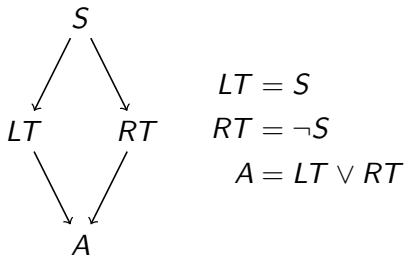
Comparing the two models, Halpern and Pearl (2005, p. 872) write:

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.

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

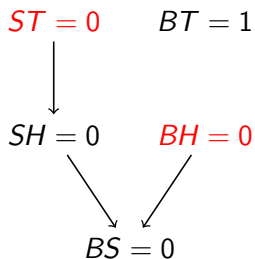
Figure: Two models of the switching scenario

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

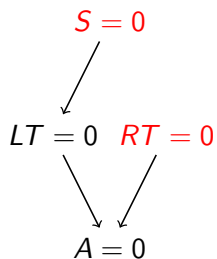
$$S = 0, LT = 0 \text{ and } RT = 0.$$

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

The two-variable model

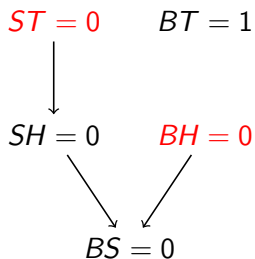


(a) Witness to Suzy causing the window to break

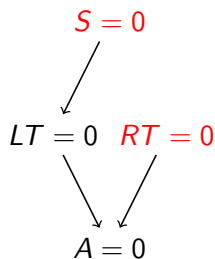


(b) Witness to the switch causing the train to arrive

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.

Questions

Is Halpern's solution to the Billy and Suzy case too sensitive to the choice of model?

Which solution do you prefer? Using production + Sartorio's Principle, or Halpern's solution?

Plan

- 1 Structural causal models
- 2 Comparison with Halpern's *Actual Causality*
- 3 **Overdetermination via fragility**
 - Cases where the effect would still have happened at the same time
- 4 Only because

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)

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?

The event fragility strategy conflicts with sufficiency

(1) 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 sufficiency condition.

Is the event fragility strategy too vague?

We would like clear predictions for clear judgements.

Are

- (2) a. Suzy throwing her rock caused the bottle to break.
- b. The enginner 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:

- (3)
 - a. Socrates died because the executioner's boat was destroyed.
 - b. The fact that the executioner's boat was destroyed caused Socrates to die.

- (4)
 - 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:

- (5) 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)

Schaffer's Merlin and Morgana scenario

Imagine that it is a law of magic that the first spell cast on a given day match the enchantment that midnight. Suppose that at noon Merlin casts a spell (the first that day) to turn the prince into a frog, that at 6:00pm Morgana casts a spell (the only other that day) to turn the prince into a frog, and that at midnight the prince becomes a frog. Clearly, Merlin's spell (the first that day) is a cause of the prince's becoming a frog and Morgana's is not, because the laws say that the first spells are the consequential ones.

(Schaffer 2000, p. 165)

Yablo's Smart Rock scenario

Billy throws a Smart Rock, equipped with an onboard computer, exquisitely designed sensors, a lightning-fast propulsion system – and instructions to make sure that the bottle shatters in exactly the way it does, at exactly the time it does. In fact, the Smart Rock doesn't need to intervene, since Suzy's throw is just right. But had it been any different – indeed, had her rock's trajectory differed in the slightest, at any point – the Smart Rock would have swooped in to make sure the job was done properly.

(Hall 2004, due to Yablo, p.c.)

- (6)
 - a. The bottle broke because Suzy threw her rock at it.
 - b. Suzy throwing her rock at the bottle caused it to break.

- (7)
 - a. The bottle broke because Billy threw his rock at it.
 - b. Billy throwing his rock at the bottle caused it to break.

Plan

- 1 Structural causal models
- 2 Comparison with Halpern's *Actual Causality*
- 3 Overdetermination via fragility
 - Cases where the effect would still have happened at the same time
- 4 Only because

- (8)
- 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.

- (8)
 - 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.

- (9)
 - 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.

- (10)
- 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.

- (10)
- 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.
- (11)
- 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 is that he drank poison.

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

- (12) **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).

- (12) **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.
- (13) 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).

(12) **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.

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

In (13a), *only* negates alternatives of the form *I introduced x to Sue*, saying I didn't introduce anyone but Bill to Sue, while in (13b) 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

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

may have a single alternative:

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

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

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

- (16) Reyna only received a Danish passport because her mother was born in Copenhagen.
- (17) Reyna received a Danish passport because her mother was born in Denmark.
- (18) does not entail (19).
 - In a world where only those born in Copenhagen receive Danish passports, (18) is true but (19) is false.
 - (In that world (19) fails the sufficiency requirement.)

Given this set of alternatives,

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

asserts that

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

is false.

Given this set of alternatives,

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

asserts that

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

is false.

$$\neg(E \text{ because } D)$$

$$\Leftrightarrow \neg\left(D \wedge (D \gg (D \text{ produce } E)) \wedge \neg(\neg D \gg (\neg D \text{ produce } E))\right)$$

$$\Leftrightarrow \neg D \vee \neg(D \gg (D \text{ produce } E)) \vee (\neg D \gg (\neg D \text{ produce } E))$$

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.

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

(21) $\Rightarrow \neg(\text{The bottle broke because Suzy or Billy threw a rock at it.})$

$\neg(\textit{Suzy or Billy throw})$

$\vee \neg((\textit{Suzy or Billy throw}) \gg ((\textit{Suzy or Billy throw}) \textit{ produce bottle break}))$

$\vee (\neg(\textit{Suzy or Billy throw}) \gg (\neg(\textit{Suzy or Billy throw}) \textit{ produce bottle break}))$





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 this account of why (20) is unacceptable.

References I

-  Hall, Ned (2004). Two concepts of causation. *Causation and counterfactuals*. Ed. by John Collins, Ned Hall, and Paul Laurie. MIT Press, pp. 225–276.
-  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.
-  Lewis, David (2000). Causation as Influence. *Journal of Philosophy* 97.4, pp. 182–197. DOI: 10.2307/2678389.

References II

-  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://www3.commons.cornell.edu/bitstream/handle/1813/28568/Rooth-1985-PhD.pdf&usg=AOvVaw2X_mAUuPyshpTVYY-Q3mK_](https://www3.commons.cornell.edu/bitstream/handle/1813/28568/Rooth-1985-PhD.pdf&usg=AOvVaw2X_mAUuPyshpTVYY-Q3mK_).
-  Schaffer, Jonathan (2000). Trumping Preemption. *Journal of Philosophy* 97.4, pp. 165–181. DOI: 10.2307/2678388.