Responses to Ahmed / Triviality

Conditionals in Decision Theory
Lecture 4

Calum McNamara and Paolo Santorio
IU and UMD

ESSLLI 2026, Prague

Recap

The story so far...

So far, we've heard about the fact that we use conditionals in deliberation.

The story so far...

So far, we've heard about the fact that we use conditionals in deliberation.

But (crudely), conditionals come in two varieties: indicatives and subjunctives.

The story so far...

So far, we've heard about the fact that we use conditionals in deliberation.

But (crudely), conditionals come in two varieties: indicatives and subjunctives.

We've argued, from two different angles, that the conditionals of deliberation are subjunctives:

- First, via applications in decision theory (causal decision theory).

The story so far...

So far, we've heard about the fact that we use conditionals in deliberation.

But (crudely), conditionals come in two varieties: indicatives and subjunctives.

We've argued, from two different angles, that the conditionals of deliberation are subjunctives:

- First, via applications in decision theory (causal decision theory).
- Second, via natural language considerations (contra DeRose).

The story so far...

So far, we've heard about the fact that we use conditionals in deliberation.

But (crudely), conditionals come in two varieties: indicatives and subjunctives.

We've argued, from two different angles, that the conditionals of deliberation are subjunctives:

- First, via applications in decision theory (causal decision theory).
- Second, via natural language considerations (contra DeRose).

Causal decision theory

The formal theory of decision-making we've focused on is causal decision theory—or at least, one version thereof.

Causal decision theory

The formal theory of decision-making we've focused on is causal decision theory—or at least, one version thereof.

This theory appeals, explicitly, to subjunctive conditionals.

Causal decision theory

The formal theory of decision-making we've focused on is causal decision theory—or at least, one version thereof.

This theory appeals, explicitly, to subjunctive conditionals.

We assumed CDT uses subjunctives that obey a selection semantics, and Lewis's miracles account of similarity.

The formal theory of decision-making we've focused on is causal decision theory—or at least, one version thereof.

This theory appeals, explicitly, to subjunctive conditionals.

We assumed CDT uses subjunctives that obey a selection semantics, and Lewis's miracles account of similarity.

And that combination seemed to get things right in classic decision problems, like Newcomb.

Causal decision theory

The formal theory of decision-making we've focused on is causal decision theory—or at least, one version thereof.

This theory appeals, explicitly, to subjunctive conditionals.

We assumed CDT uses subjunctives that obey a selection semantics, and Lewis's miracles account of similarity.

And that combination seemed to get things right in classic decision problems, like Newcomb.

But...

Ahmed's revenge

...CDT goes disastrously wrong in Ahmed's *Betting on the Laws* case.

Ahmed's revenge

...CDT goes disastrously wrong in Ahmed's *Betting on the Laws* case.

Recall the structure of that problem:

Ahmed's revenge

...CDT goes disastrously wrong in Ahmed's *Betting on the Laws* case.

Recall the structure of that problem:

- There's a proposition L, describing the (deterministic) laws of nature.

Ahmed's revenge

...CDT goes disastrously wrong in Ahmed's *Betting on the Laws* case.

Recall the structure of that problem:

- There's a proposition L, describing the (deterministic) laws of nature.
- You're more-or-less certain that L is true.

Ahmed's revenge

...CDT goes disastrously wrong in Ahmed's *Betting on the Laws* case.

Recall the structure of that problem:

- There's a proposition L, describing the (deterministic) laws of nature.
- You're more-or-less certain that L is true.
- You're certain that nothing you do can affect L's truth-value.

Ahmed's revenge

...CDT goes disastrously wrong in Ahmed's *Betting on the Laws* case.

Recall the structure of that problem:

- There's a proposition L, describing the (deterministic) laws of nature.
- You're more-or-less certain that L is true.
- You're certain that nothing you do can affect L's truth-value.
- You're then offered two bets on L, and you must choose one:

...CDT goes disastrously wrong in Ahmed's *Betting on the Laws* case.

Recall the structure of that problem:

- There's a proposition L, describing the (deterministic) laws of nature.
- You're more-or-less certain that L is true.
- You're certain that nothing you do can affect L's truth-value.
- You're then offered two bets on L, and you must choose one:
  - Bet 1: Pays \$1 if L is true, \$0 if L is false.

Ahmed's revenge

...CDT goes disastrously wrong in Ahmed's *Betting on the Laws* case.

Recall the structure of that problem:

- There's a proposition L, describing the (deterministic) laws of nature.
- You're more-or-less certain that L is true.
- You're certain that nothing you do can affect L's truth-value.
- You're then offered two bets on L, and you must choose one:
    - Bet 1: Pays \$1 if L is true, \$0 if L is false.
    - Bet 2: Pays \$0 if L is true, \$1 if L is false.

Ahmed's revenge

...CDT goes disastrously wrong in Ahmed's *Betting on the Laws* case.

Recall the structure of that problem:

- There's a proposition L, describing the (deterministic) laws of nature.
- You're more-or-less certain that L is true.
- You're certain that nothing you do can affect L's truth-value.
- You're then offered two bets on L, and you must choose one:
  - Bet 1: Pays \$1 if L is true, \$0 if L is false.
  - Bet 2: Pays \$0 if L is true, \$1 if L is false.
- Intuitively, you should choose Bet 1.

Ahmed's revenge

...CDT goes disastrously wrong in Ahmed's *Betting on the Laws* case.

Recall the structure of that problem:

- There's a proposition L, describing the (deterministic) laws of nature.
- You're more-or-less certain that L is true.
- You're certain that nothing you do can affect L's truth-value.
- You're then offered two bets on L, and you must choose one:
  - Bet 1: Pays \$1 if L is true, \$0 if L is false.
  - Bet 2: Pays \$0 if L is true, \$1 if L is false.
- Intuitively, you should choose Bet 1.
- But CDT—as we've formulated it—recommends Bet 2.
- Counterexample!

Goodbye miracles?

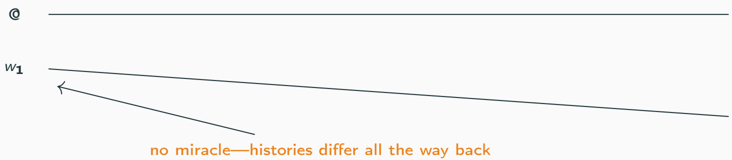
But wait! Earlier in the course, we heard that the miracles account was only one option concerning similarity—there was another.

In particular, we heard about an alternative account of similarity—one which says that the similar A-world to w is a world with identical laws, but with a history that is different from the very beginning:

Goodbye miracles?

But wait! Earlier in the course, we heard that the miracles account was only one option concerning similarity—there was another.

In particular, we heard about an alternative account of similarity—one which says that the similar A-world to w is a world with identical laws, but with a history that is different from the very beginning:



Maybe, then, what Ahmed's case teaches us is (merely) that the *miracles* account isn't the right theory of similarity, and that the alternative "different history" account is superior?

Backtracking unlimited?

Maybe, then, what Ahmed's case teaches us is (merely) that the *miracles* account isn't the right theory of similarity, and that the alternative "different history" account is superior?

Does that save us?

No refuge

No. Consider the following case (Williamson and Sandgren, 2023):

No refuge

No. Consider the following case (Williamson and Sandgren, 2023):

> *Betting on History. Let H be a proposition about the complete state of the world in the distant past. Suppose that you're almost certain that H is true. Now, suppose I offer you the following bets...*

No refuge

No. Consider the following case (Williamson and Sandgren, 2023):

> *Betting on History. Let H be a proposition about the complete state of the world in the distant past. Suppose that you're almost certain that H is true. Now, suppose I offer you the following bets...*

	H	$\neg H$
Bet 1	\$1	\$0
Bet 2	\$0	\$1

No refuge

No. Consider the following case (Williamson and Sandgren, 2023):

> *Betting on History. Let H be a proposition about the complete state of the world in the distant past. Suppose that you're almost certain that H is true. Now, suppose I offer you the following bets...*

	H	$\neg H$
Bet 1	\$1	\$0
Bet 2	\$0	\$1

Problem: We can generate exactly the same kind of problem for CDT, if we assume counterfactuals obey the "different history" account of similarity.

Side bar

Side note: maybe this version of Ahmed's case is less worrying, because it's *so* unrealistic that someone could have such high credence in a proposition, H, about the exact microscopic state of the world in the distant past.

Side bar

Side note: maybe this version of Ahmed's case is less worrying, because it's *so* unrealistic that someone could have such high credence in a proposition, H, about the exact microscopic state of the world in the distant past.

This misses the point:

Side bar

Side note: maybe this version of Ahmed's case is less worrying, because it's *so* unrealistic that someone could have such high credence in a proposition, H, about the exact microscopic state of the world in the distant past.

This misses the point:

> *Might someone respond that Betting on History is less worrying because it is more far-fetched than Betting on the Laws? We think not. CDT still advises you to do something irrational in Betting on History, and the correct decision theory never recommends the irrational. Moreover, most causalists are already motivated by the desire to have a decision theory that can handle far-fetched cases (Newcomb's problem is itself quite far-fetched). Since there is nothing incoherent about Betting on History, it falls within the purview of decision theory and we ought to be able to say something sensible about it. (Williamson and Sandgren, 2023)*

Where to go?

How, then, should we respond to Ahmed's case?

Where to go?

How, then, should we respond to Ahmed's case?

We will consider two classes of options:

Where to go?

How, then, should we respond to Ahmed's case?

We will consider two classes of options:

- Option 1: Alter the decision theory.

Where to go?

How, then, should we respond to Ahmed's case?

We will consider two classes of options:

- Option 1: Alter the decision theory.
- Option 2: Alter our theory of counterfactuals.

Alternative Decision Theories

Causal decision theory, revised

At the very start of the course, we said that there are different versions of CDT—not just Stalnaker's/Gibbard and Harper's version.

Causal decision theory, revised

At the very start of the course, we said that there are different versions of CDT—not just Stalnaker's/Gibbard and Harper's version.

Perhaps, then, what Ahmed's case teaches is merely that this *version* of CDT is inadequate:

Causal decision theory, revised

At the very start of the course, we said that there are different versions of CDT—not just Stalnaker's/Gibbard and Harper's version.

Perhaps, then, what Ahmed's case teaches is merely that this *version* of CDT is inadequate:

> *[Maybe] my example [does] not refute Causal Decision Theory as such, but only Gibbard and Harper's possibly imprudent formulation of it. (Ahmed, 2013)*

Lewisian CDT

Let us, then, briefly look at alternative version of CDT, developed by Lewis himself (1981).

The key idea is to appeal to propositions called *causal dependency hypotheses*. A causal dependency hypothesis, Lewis says, is a

Let us, then, briefly look at alternative version of CDT, developed by Lewis himself (1981).

The key idea is to appeal to propositions called *causal dependency hypotheses*. A causal dependency hypothesis, Lewis says, is a

> *a maximally specific proposition about how the things [the agent] cares about* ***do and do not depend causally*** *on his present actions.... Since there must be some truth or other on the subject, and since the dependency hypotheses are maximally specific and cannot differ without conflicting, they. comprise a partition. Exactly one of them holds at any world, and it specifies the relevant relations of causal dependence that prevail there.*

A revised equation for CDT

Let us denote the dependency hypotheses (for a given agent, at a given time) by $K_1, ..., K_n$.

Then:

Causal expected utility (Lewis's version)

$$\mathrm{CEU}(O) = \sum_i p(K_i) \cdot u(O \cap K_i)$$

Notice how similar this equation looks to our original equation for expected utility theory...

Problem solved?

If we move to Lewis's version of CDT, do we get around Ahmed's problem?

[1] The same turns out to hold for *all* other versions of CDT.

Problem solved?

If we move to Lewis's version of CDT, do we get around Ahmed's problem?

No. Suppose a dependency hypothesis, K, held fixed *both* the (deterministic) laws and the history of the world up to the time of your action, in Ahmed's original example.

[1] The same turns out to hold for *all* other versions of CDT.

Problem solved?

If we move to Lewis's version of CDT, do we get around Ahmed's problem?

No. Suppose a dependency hypothesis, K, held fixed *both* the (deterministic) laws and the history of the world up to the time of your action, in Ahmed's original example.

Then, we would have a situation like the following:

	$H_1 \wedge L$	$H_2 \wedge L$	...
Bet 1	\$1	$\varnothing$	...
Bet 2	$\varnothing$	\$0	...

[1] The same turns out to hold for *all* other versions of CDT.

Problem solved?

If we move to Lewis's version of CDT, do we get around Ahmed's problem?

No. Suppose a dependency hypothesis, K, held fixed *both* the (deterministic) laws and the history of the world up to the time of your action, in Ahmed's original example.

Then, we would have a situation like the following:

	$H_1 \wedge L$	$H_2 \wedge L$	...
Bet 1	\$1	$\varnothing$	...
Bet 2	$\varnothing$	\$0	...

This is not even a well-formed decision problem. So dependency hypotheses cannot hold fixed *everything* that's causally independent of your choice either.[1]

[1] The same turns out to hold for *all* other versions of CDT.

More theories

Another option, then, is to endorse theories that require more extreme changes to causal decision theory. Here are some options:

- Sandgren and Williamson (2020) propose a version of CDT—*selective causal decision theory*—in which you simply ignore outcomes that you think are inconsistent with the laws.

More theories

Another option, then, is to endorse theories that require more extreme changes to causal decision theory. Here are some options:

- Sandgren and Williamson (2020) propose a version of CDT—*selective causal decision theory*—in which you simply ignore outcomes that you think are inconsistent with the laws.
- Kment (2023) proposes a decision theory very close to EDT, in which you conditionalize, not on the proposition that some action is performed, but on the proposition that the action is *possible*:

More theories

Another option, then, is to endorse theories that require more extreme changes to causal decision theory. Here are some options:

- Sandgren and Williamson (2020) propose a version of CDT—*selective causal decision theory*—in which you simply ignore outcomes that you think are inconsistent with the laws.
- Kment (2023) proposes a decision theory very close to EDT, in which you conditionalize, not on the proposition that some action is performed, but on the proposition that the action is *possible*:

$$\mathrm{KEU}(O) = \sum_i p(S_i \mid \Diamond A) \cdot u(A \wedge S_i)$$

More theories

Another option, then, is to endorse theories that require more extreme changes to causal decision theory. Here are some options:

- Sandgren and Williamson (2020) propose a version of CDT—*selective causal decision theory*—in which you simply ignore outcomes that you think are inconsistent with the laws.
- Kment (2023) proposes a decision theory very close to EDT, in which you conditionalize, not on the proposition that some action is performed, but on the proposition that the action is *possible*:

$$\mathrm{KEU}(O) = \sum_i p(S_i \mid \Diamond A) \cdot u(A \wedge S_i)$$

- Fusco (2023) argues that Ahmed's problems can be avoided if we appeal to a process of *dynamic deliberation*.

Problems, problems

However, none of these theories is satisfactory (by our lights):

Problems, problems

However, none of these theories is satisfactory (by our lights):

- Sandgren and Williamson's theory simply does not give the right answer in the *Betting on the Laws* case (although it gives the right answer in closely related cases).

Problems, problems

However, none of these theories is satisfactory (by our lights):

- Sandgren and Williamson's theory simply does not give the right answer in the *Betting on the Laws* case (although it gives the right answer in closely related cases).
- Kment's theory requires one-boxing in the Newcomb problem (at least some of the time, depending on your exact probabilities).

Problems, problems

However, none of these theories is satisfactory (by our lights):

- Sandgren and Williamson's theory simply does not give the right answer in the *Betting on the Laws* case (although it gives the right answer in closely related cases).
- Kment's theory requires one-boxing in the Newcomb problem (at least some of the time, depending on your exact probabilities).
- Fusco's solution requires you to be able to conditionalize your credences on inconsistent propositions, like $H \wedge L$.

Joyce's (dis)solution

Joyce (2016) responds to Ahmed's cases in a very different way:

Joyce (2016) responds to Ahmed's cases in a very different way:

> *an agent who deliberates about a decision which is framed [like Ahmed's] is engaging in an epistemic exercise, not an agential one.*

Joyce (2016) responds to Ahmed's cases in a very different way:

> *an agent who deliberates about a decision which is framed [like Ahmed's] is engaging in an epistemic exercise, not an agential one.*

Broadly, then: Joyce thinks that cases like Ahmed's—where the possibility of determinism is salient—are not real decision problems.

Ahmed himself entertains this idea:

Incompatibilism

Ahmed himself entertains this idea:

> *One way to hold on to Causal Decision Theory... is to invoke the Kantian line that rational deliberation (that is, of an agent) must proceed on an assumption of indeterminism. If Kant was right about that, and if your assuming indeterminism entails that your $p(L) < 0.5$ (which is surely the least that it requires), then there is no difficulty in asserting... CDT....*

Ahmed himself entertains this idea:

> *One way to hold on to Causal Decision Theory... is to invoke the Kantian line that rational deliberation (that is, of an agent) must proceed on an assumption of indeterminism. If Kant was right about that, and if your assuming indeterminism entails that your $p(L) < 0.5$ (which is surely the least that it requires), then there is no difficulty in asserting... CDT....*

This is a kind of incompatibilism about free will. We would strongly like to avoid this response.

Semantic solutions

A different style of solution: change the semantics of counterfactuals.

Semantic solutions

A different style of solution: change the semantics of counterfactuals.

A diagnosis of the Ahmed problem:

- The key issue is independent of decision theory, and lies in the way we define similarity for subjunctives.

Semantic solutions

A different style of solution: change the semantics of counterfactuals.

A diagnosis of the Ahmed problem:

- The key issue is independent of decision theory, and lies in the way we define similarity for subjunctives.
- In particular, recall that Ahmed exploits the following entailment (where 'M' stands for the miracles view of similarity):

$$Bet1 \mathrel{\Box\!\!\rightarrow} L \wedge M \vDash Bet2 \mathrel{\Box\!\!\rightarrow} \neg L$$

Semantic solutions

A different style of solution: change the semantics of counterfactuals.

A diagnosis of the Ahmed problem:

- The key issue is independent of decision theory, and lies in the way we define similarity for subjunctives.
- In particular, recall that Ahmed exploits the following entailment (where 'M' stands for the miracles view of similarity):

$$Bet1 \mathrel{\Box\!\!\rightarrow} L \wedge M \models Bet2 \mathrel{\Box\!\!\rightarrow} \neg L$$

- This entailment seems dubious, especially if you find conditionals of the following form suspicious:

(4) If I had scratched my head once more yesterday, the laws would have been different.

Semantic solutions

We will see two semantic solutions to the problem:

(i) Williamson and Sangren on impossible worlds (2021);
(ii) Gallow on causal models and failures of agglomeration.

Williamson and Sandgren (2021): impossible worlds

As we've seen, standard modal semantics exploit a space of possible worlds.

Williamson and Sandgren (2021): impossible worlds

As we've seen, standard modal semantics exploit a space of possible worlds.

But: some theorists also want to admit impossible worlds.

Williamson and Sandgren (2021): impossible worlds

As we've seen, standard modal semantics exploit a space of possible worlds.

But: some theorists also want to admit impossible worlds.

What are impossible worlds?

- Think of possible worlds as maximal consistent descriptions of a state of affairs (linguistic ersatzism).

Williamson and Sandgren (2021): impossible worlds

As we've seen, standard modal semantics exploit a space of possible worlds.

But: some theorists also want to admit impossible worlds.

What are impossible worlds?

- Think of possible worlds as maximal consistent descriptions of a state of affairs (linguistic ersatzism).
- Impossible worlds are just like that, except... they allow for localized inconsistencies.

Williamson and Sandgren (2021): impossible worlds

As we've seen, standard modal semantics exploit a space of possible worlds.

But: some theorists also want to admit impossible worlds.

What are impossible worlds?

- Think of possible worlds as maximal consistent descriptions of a state of affairs (linguistic ersatzism).
- Impossible worlds are just like that, except... they allow for localized inconsistencies.
- Accepting impossible worlds has been an option for a long time, though a minoritary one (see especially Nolan 1997).

Williamson and Sandgren (2021): impossible worlds

How do impossible worlds help?

Williamson and Sandgren (2021): impossible worlds

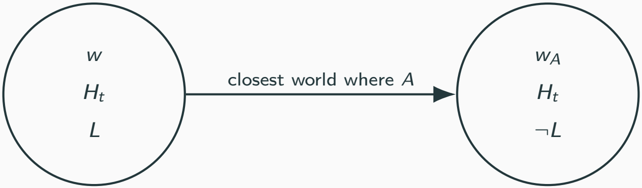
How do impossible worlds help?

- Williamson and Sandgren observe: miracle accounts make laws of nature modally fragile.
- Make a counterfactual supposition that takes you to a different world, and you will change the laws.

Williamson and Sandgren (2021): impossible worlds

How do impossible worlds help?

- Williamson and Sandgren observe: miracle accounts make laws of nature modally fragile.
- Make a counterfactual supposition that takes you to a different world, and you will change the laws.



Williamson and Sandgren (2021): impossible worlds

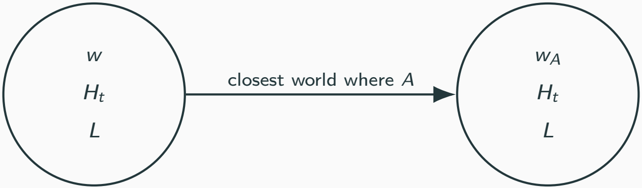
The suggestion: we can restore modal robustness of the laws by adopting impossible worlds.

- In Lewis's semantics, closest worlds share a history, up to a time.
- On the assumption of determinism, this forces us to say that they don't share laws (since some facts are different).
- ...but we can say that closest worlds share both a history and laws, and differ in some matter of facts, if worlds are allowed to contain inconsistencies.

Williamson and Sandgren (2021): impossible worlds

The suggestion: we can restore modal robustness of the laws by adopting impossible worlds.

- In Lewis's semantics, closest worlds share a history, up to a time.
- On the assumption of determinism, this forces us to say that they don't share laws (since some facts are different).
- ...but we can say that closest worlds share both a history and laws, and differ in some matter of facts, if worlds are allowed to contain inconsistencies.



Williamson and Sandgren (2021): worries

Two worries about the impossible-worlds strategy:

Williamson and Sandgren (2021): worries

Two worries about the impossible-worlds strategy:

- Baby/bathwater. The original problem was that possible-world similarity makes laws/history too fragile. Moving to impossible worlds changes the modal space itself.

Williamson and Sandgren (2021): worries

Two worries about the impossible-worlds strategy:

- Baby/bathwater. The original problem was that possible-world similarity makes laws/history too fragile. Moving to impossible worlds changes the modal space itself.

- Stipulation worry. The solution needs the closest A-world to preserve exactly the features we wanted held fixed: H_t and L. But why is *that* impossible world closest?

Williamson and Sandgren (2021): worries

Two worries about the impossible-worlds strategy:

- Baby/bathwater. The original problem was that possible-world similarity makes laws/history too fragile. Moving to impossible worlds changes the modal space itself.
- Stipulation worry. The solution needs the closest A-world to preserve exactly the features we wanted held fixed: H_t and L. But why is *that* impossible world closest?

Gallow (2022): a causal model

Gallow evaluates counterfactuals by intervening on a causal model, rather than by consulting the closest world directly.

Gallow (2022): a causal model

Gallow evaluates counterfactuals by intervening on a causal model, rather than by consulting the closest world directly.

For *Betting on the Laws*, let:

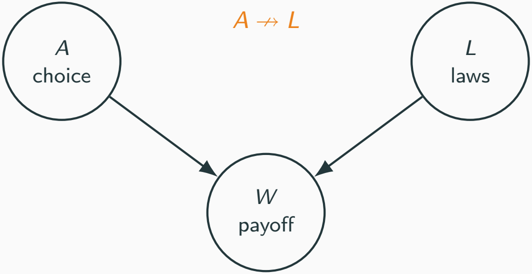
$$A = \text{you take Bet 2}, \qquad L = \text{the laws hold}, \qquad W = \text{your payoff}.$$

Gallow (2022): a causal model

Gallow evaluates counterfactuals by intervening on a causal model, rather than by consulting the closest world directly.

For *Betting on the Laws*, let:

$A =$ you take Bet 2, $\quad L =$ the laws hold, $\quad W =$ your payoff.

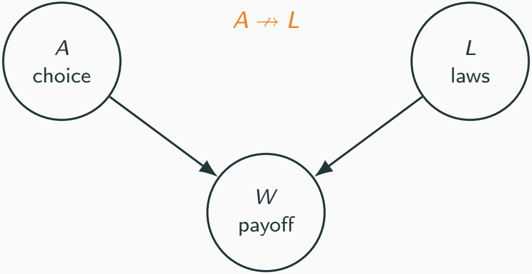


Gallow evaluates counterfactuals by intervening on a causal model, rather than by consulting the closest world directly.

For *Betting on the Laws*, let:

$$A = \text{you take Bet 2}, \qquad L = \text{the laws hold}, \qquad W = \text{your payoff}.$$

$A \nrightarrow L$



Choice and the laws both affect the payoff; but your choice does not causally affect whether the laws hold.

Gallow (2022): blocking Ahmed

Suppose Bet 1 is actual and L is true. To ask what would happen under Bet 2, intervene by setting A to true.

Gallow (2022): blocking Ahmed

Suppose Bet 1 is actual and L is true. To ask what would happen under Bet 2, intervene by setting A to true.

No arrow from A to L $\implies$ L remains fixed.

Gallow (2022): blocking Ahmed

Suppose Bet 1 is actual and L is true. To ask what would happen under Bet 2, intervene by setting A to true.

$$\text{No arrow from } A \text{ to } L \quad \Longrightarrow \quad L \text{ remains fixed.}$$

Hence:

$$A \mathrel{\Box\!\!\rightarrow} L.$$

Gallow (2022): blocking Ahmed

Suppose Bet 1 is actual and L is true. To ask what would happen under Bet 2, intervene by setting A to true.

$$\text{No arrow from } A \text{ to } L \quad \Longrightarrow \quad L \text{ remains fixed.}$$

Hence:

$$A \mathrel{\Box\!\!\rightarrow} L.$$

So Bet 2 would pay \$0, not \$1; Bet 1 pays \$1. CDT therefore recommends Bet 1.

Suppose Bet 1 is actual and L is true. To ask what would happen under Bet 2, intervene by setting A to true.

No arrow from A to L $\implies$ L remains fixed.

Hence:

$$A \mathrel{\Box\!\!\rightarrow} L.$$

So Bet 2 would pay \$0, not \$1; Bet 1 pays \$1. CDT therefore recommends Bet 1.

The same model structure gives $A \mathrel{\Box\!\!\rightarrow} H$ in *Betting on History*, where H describes the actual past.

Gallow (2022): no Agglomeration

Gallow can validate the two counterfactuals separately:

$$A \mathrel{\Box\!\!\rightarrow} L \qquad\qquad A \mathrel{\Box\!\!\rightarrow} H.$$

Gallow (2022): no Agglomeration

Gallow can validate the two counterfactuals separately:

$$A \mathrel{\Box\!\!\rightarrow} L \qquad\qquad A \mathrel{\Box\!\!\rightarrow} H.$$

But they do not agglomerate:

$$A \mathrel{\Box\!\!\rightarrow} (L \wedge H)$$

does not follow.

Gallow (2022): no Agglomeration

Gallow can validate the two counterfactuals separately:

$$A \mathrel{\Box\!\!\rightarrow} L \qquad\qquad A \mathrel{\Box\!\!\rightarrow} H.$$

But they do not agglomerate:

$$A \mathrel{\Box\!\!\rightarrow} (L \wedge H)$$

does not follow.

Under determinism, the actual laws and actual past determine the actual action:

$$L \wedge H \vDash \neg A.$$

Gallow (2022): no Agglomeration

Gallow can validate the two counterfactuals separately:

$$A \mathrel{\Box\!\!\rightarrow} L \qquad\qquad A \mathrel{\Box\!\!\rightarrow} H.$$

But they do not agglomerate:

$$A \mathrel{\Box\!\!\rightarrow} (L \wedge H)$$

does not follow.

Under determinism, the actual laws and actual past determine the actual action:

$$L \wedge H \vDash \neg A.$$

So A, L, and H cannot be independently varied. Gallow's distinctness requirement therefore rules out a single causal model containing all three.

Gallow (2022): the final case

This leads to problems. Let M be ‘there are no miracles’, and let C be ‘The laws remain unchanged’.

Now consider:

	$M \wedge C$	$\neg(M \wedge C)$
Bet 1	\$1	\$0
Bet 2	\$0	\$1

Gallow (2022): the final case

This leads to problems. Let M be 'there are no miracles', and let C be 'The laws remain unchanged'.

Now consider:

	$M \wedge C$	$\neg(M \wedge C)$
Bet 1	\$1	\$0
Bet 2	\$0	\$1

CDT now needs to know which of these is true:

$$A \mathrel{\Box\!\!\rightarrow} (L \wedge H)$$

$$A \mathrel{\Box\!\!\rightarrow} \neg(L \wedge H)$$

Gallow (2022): the final case

This leads to problems. Let M be 'there are no miracles', and let C be 'The laws remain unchanged'.

Now consider:

	$M \wedge C$	$\neg(M \wedge C)$
Bet 1	\$1	\$0
Bet 2	\$0	\$1

CDT now needs to know which of these is true:

$$A \mathrel{\Box\!\!\rightarrow} (L \wedge H)$$
$$A \mathrel{\Box\!\!\rightarrow} \neg(L \wedge H)$$

Problem: the joint causal model, containing an "action variable", M, and C is inadmissible. So, Gallow's semantics is silent. The proposal handles the earlier cases separately, but not something like their combination in full generality.

Triviality and decision, 1/2: indicatives

Triviality and decision

- The key idea we explored in this class: conditionals play a key role in evaluating decision.
- For this to work: the probabilities of conditionals (both epistemic and causal) have to conform to some (pretty intuitive) principles.

Triviality and decision

- The key idea we explored in this class: conditionals play a key role in evaluating decision.
- For this to work: the probabilities of conditionals (both epistemic and causal) have to conform to some (pretty intuitive) principles.

But: there are arguments to the effect that probabilities of conditionals are not as well-behaved.

Triviality and decision

A quote we saw in lecture 1:

> *Some think that in (a suitable version of) Newcomb's problem, it is rational to take only one box. These one-boxers think of the situation as a choice between a million and a thousand. They [viz., evidential decision theorists]* ***are convinced by indicative conditionals: if I take one box I will be a millionaire, but if I take both boxes I will not****.*

Triviality and decision

A quote we saw in lecture 1:

> *Some think that in (a suitable version of) Newcomb's problem, it is rational to take only one box. These one-boxers think of the situation as a choice between a million and a thousand. They [viz., evidential decision theorists]* ***are convincedby indicative conditionals: if I take one box I will be a millionaire, but if I take both boxes I will not****.*

Here Lewis seems to be assuming the equivalence of the following two:

(i) Evidential EU of O: $\sum_i p(S_i \mid O) \cdot u(O \cap S_i)$

(ii) Evidential EU of O: $\sum_i p(O \to S_i) \cdot u(O \cap S_i)$

But this seems to presuppose the general validity of a controversial principle...

A fair six-sided die was tossed.

Stalnaker's Thesis

A fair six-sided die was tossed.

Say that a die lands low if it lands on 1, 2, or 3, and that it lands high if it lands on 4, 5, or 6.

Stalnaker's Thesis

A fair six-sided die was tossed.

Say that a die lands low if it lands on 1, 2, or 3, and that it lands high if it lands on 4, 5, or 6.

What credence would you assign to (5)?

(7) If the die landed low, it landed on 2.

A fair six-sided die was tossed.

Say that a die lands low if it lands on 1, 2, or 3, and that it lands high if it lands on 4, 5, or 6.

What credence would you assign to (5)?

(8) If the die landed low, it landed on 2.

The intuitive answer, of course: 1/3.

Stalnaker's Thesis

Besides being intuitive, this answer is supported by so-called Stalnaker's Thesis.

Stalnaker's Thesis. For all rational $Pr : Pr(\mathrm{A}) > 0$, all A, B:

$$Pr(\mathrm{A} \to \mathrm{B}) = Pr(\mathrm{B} \mid \mathrm{A})$$

Informally: ST says that the probability of a conditional $\mathrm{A} \to \mathrm{B}$ should line up with the conditional probability $Pr(\mathrm{B} \mid \mathrm{A})$.

Bridge principles

Stalnaker's Thesis (ST). For all rational Pr : $Pr(\mathrm{A}) > 0$, all A, B:

$$Pr(\mathrm{A} \to \mathrm{B}) = Pr(\mathrm{B} \mid \mathrm{A})$$

Why believe Stalnaker's Thesis?

Bridge principles

Stalnaker's Thesis (ST). For all rational Pr : $Pr(A) > 0$, all A, B:

$$Pr(A \to B) = Pr(B \mid A)$$

Why believe Stalnaker's Thesis?

(i) ST systematically captures intuitive probability judgments.

Bridge principles

Stalnaker's Thesis (ST). For all rational Pr : $Pr(\mathsf{A}) > 0$, all A, B:

$$Pr(\mathsf{A} \to \mathsf{B}) = Pr(\mathsf{B} \mid \mathsf{A})$$

Why believe Stalnaker's Thesis?

(i) ST systematically captures intuitive probability judgments.

(ii) ST provides a desirable link between conditionals and supposition.

(Assuming: conditional probabilities are probabilities under supposition.)

Bridge principles

Stalnaker's Thesis is a bridge principle: it states a connection between natural language sentences of a certain form and rational credence.

Bridge principles

Stalnaker's Thesis is a bridge principle: it states a connection between natural language sentences of a certain form and rational credence.

- ST is the most famous principle of this sort.
- But: it is by no means the only one!
- Many plausible bridge principles btw epistemic language and probability.

Stalnaker's Thesis

Unfortunately, Stalnaker's Thesis is subject to so-called triviality results.

- These results show that, by assuming ST and seemingly innocent side premises, we can infer absurd conclusions.
- Example: Lewis 1976 proves that, whenever $Pr(\mathrm{A}) > 0$, $Pr(\mathrm{A} \to \mathrm{B}) = Pr(\mathrm{B})$.

Unfortunately, Stalnaker's Thesis is subject to so-called triviality results.

- These results show that, by assuming ST and seemingly innocent side premises, we can infer absurd conclusions.
- Example: Lewis 1976 proves that, whenever $Pr(\mathrm{A}) > 0$, $Pr(\mathrm{A} \to \mathrm{B}) = Pr(\mathrm{B})$.

Moreover: triviality affects a large number of modality-probability bridge principles.

- This is not just about conditionals.
- It's about epistemic modal language in general.

A triviality proof

Another bridge principle about conditionals (Bradley 2000, 2007):

Positive Preservation (PP).
For all rational Pr, and all A, B such that $Pr(\text{A}) > 0$:

$$\text{If } Pr(\text{B}) = 1, \text{ then } Pr(\text{A} \to \text{B}) = 1$$

A triviality proof

Another bridge principle about conditionals (Bradley 2000, 2007):

Positive Preservation (PP).
For all rational Pr, and all A, B such that $Pr(\text{A}) > 0$:

$$\text{If } Pr(\text{B}) = 1, \text{ then } Pr(\text{A} \rightarrow \text{B}) = 1$$

- PP is entailed by Stalnaker's Thesis, given minimal assumptions.
- But it is also independently plausible.

Suppose you have credence 1 that the die landed low. Consider:

(10) If the die landed even, it landed low.

What credence do you have in (9)? Arguably, 1 as well.

In the rest of the lecture: a basic triviality result for conditionals.

Some basic assumptions

In the rest of the lecture: a basic triviality result for conditionals.

We will show: starting from some (seemingly) basic constraints on probability functions, we can get unacceptable conclusions.

Some basic assumptions

In the rest of the lecture: a basic triviality result for conditionals.

We will show: starting from some (seemingly) basic constraints on probability functions, we can get unacceptable conclusions.

In what follows:

- 'Pr' stands for a rational agent's credence function.
- 'Pr_{A}' stands for a rational agent's credence function, updated on A. (Update can stand for both learning and supposition.)

Some basic assumptions

We will assume the following Bayesian principles:

(i) Identity. For all Pr and all consistent A, $Pr_{\mathsf{A}}(\mathsf{A}) = 1$

(ii) Conjunction. For all rational Pr, and for all A and B: $Pr(\mathsf{A} \wedge \mathsf{B}) \leq Pr(\mathsf{A})$

(iii) Ratio. For all A and B: $Pr(\mathsf{B} \mid \mathsf{A}) = \dfrac{Pr(\mathsf{A} \wedge \mathsf{B})}{Pr(\mathsf{A})}$

(iv) Conditionalization. For all rational Pr and all A: $Pr_{\mathsf{A}}(\bullet) = Pr(\bullet \mid \mathsf{A})$

(v) Closure. If Pr is rational, then $Pr_{\mathsf{A}}(\bullet)$ is rational,
for any A such that $Pr(\mathsf{A}) > 0$.

Some basic assumptions

(i) Identity. For all Pr and all consistent A, $Pr_{\mathsf{A}}(\mathsf{A}) = 1$

Some basic assumptions

(i) Identity. For all Pr and all consistent A, $Pr_{\mathsf{A}}(\mathsf{A}) = 1$

- Identity says that, under the supposition that A, you should have credence 1 in A.
- This is accepted by any theory of learning/supposition.
- It merely says: under the supposition that A, A is certain.

Some basic assumptions

(ii) Conjunction. For all rational Pr, and for all A and B: $Pr(\mathrm{A} \wedge \mathrm{B}) \leq Pr(\mathrm{A})$

Some basic assumptions

(ii) Conjunction. For all rational Pr, and for all A and B: $Pr(\mathrm{A} \wedge \mathrm{B}) \leq Pr(\mathrm{A})$

- This principle says that the probability of a conjunction should be at most as high as the probability of a conjunct.
- It is, again, accepted by virtually any notion of porbbaility.

Some basic assumptions

(iii) Ratio. For all A and B: $Pr(\mathrm{B} \mid \mathrm{A}) = \dfrac{Pr(\mathrm{A} \wedge \mathrm{B})}{Pr(\mathrm{A})}$

Some basic assumptions

(iii) Ratio. For all A and B: $Pr(\mathrm{B} \mid \mathrm{A}) = \dfrac{Pr(\mathrm{A} \wedge \mathrm{B})}{Pr(\mathrm{A})}$

This is just the classical definition of conditional probability.

(iv) Conditionalization. For all rational Pr and all A: $Pr_{\mathsf{A}}(\bullet) = Pr(\bullet \mid \mathsf{A})$

Some basic assumptions

(iv) Conditionalization. For all rational Pr and all A: $Pr_{\mathsf{A}}(\bullet) = Pr(\bullet \mid \mathsf{A})$

- This says that probability under supposition should be identical to conditional probabilities.
- It is the standard Bayesian way to model supposition.

(v) Closure. If Pr is rational, then $Pr_{\mathsf{A}}(\bullet)$ is rational, for any A such that $Pr(\mathsf{A}) > 0$.

(v) Closure. If Pr is rational, then $Pr_{A}(\bullet)$ is rational, for any A such that $Pr(A) > 0$.

- This says that, if you start from a rational probability function and you suppose a proposition that you're not certain is false, you still get a rational probability function.
- In a slogan: if you're rational, merely supposing a proposition that you regard as possible cannot make you irrational.

A triviality proof

In addition to these, for our proof we assume:

Positive Preservation (PP). For all Pr, and all A, B s.t. $Pr(\text{A}) > 0$:

$$\text{If } Pr(\text{B}) = 1, \text{ then } Pr(\text{A} \rightarrow \text{B}) = 1$$

A triviality proof

Assume $Pr(\mathrm{A}) > 0$.

A triviality proof

Assume $Pr(\mathsf{A}) > 0$.

i. $Pr(\mathsf{A} \rightarrow \mathsf{B}) \geq Pr((\mathsf{A} \rightarrow \mathsf{B}) \wedge \mathsf{B})$ (Conjunction)

A triviality proof

Assume $Pr(\mathsf{A}) > 0$.

i. $Pr(\mathsf{A} \to \mathsf{B}) \geq Pr((\mathsf{A} \to \mathsf{B}) \wedge \mathsf{B})$ (Conjunction)

ii. $Pr(\mathsf{A} \to \mathsf{B}) \geq Pr(\mathsf{B}) \times Pr(\mathsf{A} \to \mathsf{B} \mid \mathsf{B})$ (i, Ratio)

A triviality proof

Assume $Pr(\mathsf{A}) > 0$.

i. $Pr(\mathsf{A} \rightarrow \mathsf{B}) \geq Pr((\mathsf{A} \rightarrow \mathsf{B}) \wedge \mathsf{B})$ (Conjunction)

ii. $Pr(\mathsf{A} \rightarrow \mathsf{B}) \geq Pr(\mathsf{B}) \times Pr(\mathsf{A} \rightarrow \mathsf{B} \mid \mathsf{B})$ (i, Ratio)

iii. $Pr(\mathsf{A} \rightarrow \mathsf{B}) \geq Pr(\mathsf{B}) \times Pr_{\mathsf{B}}(\mathsf{A} \rightarrow \mathsf{B})$ (ii, Conditionalization)

A triviality proof

Assume $Pr(\mathrm{A}) > 0$.

i. $Pr(\mathrm{A} \to \mathrm{B}) \geq Pr((\mathrm{A} \to \mathrm{B}) \wedge \mathrm{B})$ (Conjunction)

ii. $Pr(\mathrm{A} \to \mathrm{B}) \geq Pr(\mathrm{B}) \times Pr(\mathrm{A} \to \mathrm{B} \mid \mathrm{B})$ (i, Ratio)

iii. $Pr(\mathrm{A} \to \mathrm{B}) \geq Pr(\mathrm{B}) \times Pr_{\mathrm{B}}(\mathrm{A} \to \mathrm{B})$ (ii, Conditionalization)

iv. $Pr_{\mathrm{B}}(\mathrm{B}) = 1$ (Identity)

A triviality proof

Assume $Pr(\mathrm{A}) > 0$.

i. $Pr(\mathrm{A} \to \mathrm{B}) \geq Pr((\mathrm{A} \to \mathrm{B}) \wedge \mathrm{B})$ (Conjunction)

ii. $Pr(\mathrm{A} \to \mathrm{B}) \geq Pr(\mathrm{B}) \times Pr(\mathrm{A} \to \mathrm{B} \mid \mathrm{B})$ (i, Ratio)

iii. $Pr(\mathrm{A} \to \mathrm{B}) \geq Pr(\mathrm{B}) \times Pr_{\mathrm{B}}(\mathrm{A} \to \mathrm{B})$ (ii, Conditionalization)

iv. $Pr_{\mathrm{B}}(\mathrm{B}) = 1$ (Identity)

v. $Pr_{\mathrm{B}}(\mathrm{A} \to \mathrm{B}) = 1$ (iv, PP, Closure)

A triviality proof

Assume $Pr(\mathsf{A}) > 0$.

i. $Pr(\mathsf{A} \to \mathsf{B}) \geq Pr((\mathsf{A} \to \mathsf{B}) \wedge \mathsf{B})$ (Conjunction)

ii. $Pr(\mathsf{A} \to \mathsf{B}) \geq Pr(\mathsf{B}) \times Pr(\mathsf{A} \to \mathsf{B} \mid \mathsf{B})$ (i, Ratio)

iii. $Pr(\mathsf{A} \to \mathsf{B}) \geq Pr(\mathsf{B}) \times Pr_{\mathsf{B}}(\mathsf{A} \to \mathsf{B})$ (ii, Conditionalization)

iv. $Pr_{\mathsf{B}}(\mathsf{B}) = 1$ (Identity)

v. $Pr_{\mathsf{B}}(\mathsf{A} \to \mathsf{B}) = 1$ (iv, PP, Closure)

vi. $Pr(\mathsf{A} \to \mathsf{B}) \geq Pr(\mathsf{B}) \times 1 = Pr(\mathsf{B})$ (iii, v)

A triviality proof

We proved that, whenever $Pr(\mathrm{A}) > 0$, $Pr(\mathrm{A} \to \mathrm{B}) \geq Pr(\mathrm{B})$.

A triviality proof

We proved that, whenever $Pr(\mathrm{A}) > 0$, $Pr(\mathrm{A} \rightarrow \mathrm{B}) \geq Pr(\mathrm{B})$. This is absurd.

A triviality proof

We proved that, whenever $Pr(\mathrm{A}) > 0$, $Pr(\mathrm{A} \to \mathrm{B}) \geq Pr(\mathrm{B})$. This is absurd.

Counterexample:

(13) a. The die landed low.

b. If the die landed even, it landed low.

Intuitively: $Pr(\textit{low}) = 1/2$ and $Pr(\textit{even} \to \textit{low}) = 1/3$.

The options

So we proved an absurd result, starting from seemingly acceptable premises!

The options

So we proved an absurd result, starting from seemingly acceptable premises!

What are the options?

- Deny Preservation (and Stalnaker's Thesis!).

$\rightsquigarrow$ If so, forget about using conditionals in DT.

The options

So we proved an absurd result, starting from seemingly acceptable premises!

What are the options?

- Deny Preservation (and Stalnaker's Thesis!).

$\rightsquigarrow$ If so, forget about using conditionals in DT.

- Block triviality via context dependence (Van Fraassen 1976, Bacon 2015).

The options

So we proved an absurd result, starting from seemingly acceptable premises!

What are the options?

- Deny Preservation (and Stalnaker's Thesis!).

$\rightsquigarrow$ If so, forget about using conditionals in DT.

- Block triviality via context dependence (Van Fraassen 1976, Bacon 2015).
- Deny some classical principle about probability
  (e.g. Conditionalization—see Goldstein and Santorio 2021.)

The options

So we proved an absurd result, starting from seemingly acceptable premises!

What are the options?

- Deny Preservation (and Stalnaker's Thesis!).

$\rightsquigarrow$ If so, forget about using conditionals in DT.

- Block triviality via context dependence (Van Fraassen 1976, Bacon 2015).
- Deny some classical principle about probability
  (e.g. Conditionalization—see Goldstein and Santorio 2021.)

To be continued...

References

Arif Ahmed. Causal decision theory: A counterexample. *Philosophical Review*, 122(2): 289–306, 2013. doi: 10.1215/00318108-1963725. URL https://doi.org/10.1215%2F00318108-1963725.

Melissa Fusco. Absolution of a causal decision theorist. *Noûs*, jun 2023. doi: 10.1111/nous.12459. URL https://doi.org/10.1111%2Fnous.12459.

James M. Joyce. Review of Arif Ahmed: Evidence, Decision and Causality. *Journal of Philosophy*, 113(4):224–232, 2016. ISSN 0022-362X. doi: 10.5840/jphil2016113413. URL http://dx.doi.org/10.5840/jphil2016113413.

Boris Kment. Decision, causality, and predetermination. *Philosophy and Phenomenological Research*, jan 2023. doi: 10.1111/phpr.12935. URL https://doi.org/10.1111%2Fphpr.12935.

David Lewis. Causal decision theory. *Australasian Journal of Philosophy*, 59(1):5–30, Mar 1981. ISSN 1471-6828. doi: 10.1080/00048408112340011. URL http://dx.doi.org/10.1080/00048408112340011.

Alexander Sandgren and Timothy Luke Williamson. Determinism, counterfactuals, and decision. *Australasian Journal of Philosophy*, 99(2):286–302, jun 2020. doi: 10.1080/00048402.2020.1764073. URL https://doi.org/10.1080%2F00048402.2020.1764073.

Timothy Luke Williamson and Alexander Sandgren. Law-abiding causal decision theory. *British Journal for the Philosophy of Science*, 2023. doi: 10.1086/715103.