# Triviality All Ways

Conditionals in Decision Theory
Lecture 5

Calum McNamara and Paolo Santorio
IU and UMD

ESSLLI 2026, Prague

# Recap

# Where we are

We have seen several versions of causal decision theory.

# Where we are

We have seen several versions of causal decision theory.

They are supposed to share a common idea:

> Choose by tracking how the world causally depends on your options.

# Where we are

We have seen several versions of causal decision theory.

They are supposed to share a common idea:

> Choose by tracking how the world causally depends on your options.

But they implement that idea in different formal languages.

# Three versions of CDT

Let $O$ be an option, let $S_i$ be a partition of relevant states, and let $u(O \cap S_i)$ be the utility of the corresponding outcome.

# Three versions of CDT

Let $O$ be an option, let $S_i$ be a partition of relevant states, and let $u(O \cap S_i)$ be the utility of the corresponding outcome.

- Stalnakerian CDT

$$U_{\mathrm{ST}}(O) = \sum_i p(O \mathrel{\Box\!\!\rightarrow} S_i) \cdot u(O \cap S_i)$$

# Three versions of CDT

Let $O$ be an option, let $S_i$ be a partition of relevant states, and let $u(O \cap S_i)$ be the utility of the corresponding outcome.

- Stalnakerian CDT

$$U_{\mathrm{ST}}(O) = \sum_i p(O \mathrel{\Box\!\!\rightarrow} S_i) \cdot u(O \cap S_i)$$

- Lewisian CDT

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

where the $K_i$ are dependency hypotheses.

# Three versions of CDT

Let $O$ be an option, let $S_i$ be a partition of relevant states, and let $u(O \cap S_i)$ be the utility of the corresponding outcome.

- Stalnakerian CDT

$$U_{\mathrm{ST}}(O) = \sum_i p(O \mathrel{\Box\!\!\rightarrow} S_i) \cdot u(O \cap S_i)$$

- Lewisian CDT

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

  where the $K_i$ are dependency hypotheses.

- Skyrmsian CDT (a version we haven't seen before)

$$U_{\mathrm{SK}}(O) = \sum_i \sum_w p(w) \cdot ch_{H_{w,t^-}}(S_i \mid O) \cdot u(O \cap S_i)$$

# The equivalence thought

These theories look different.

These theories look different.

But a common thought is that the differences are merely notational, or at least merely matters of formulation.

# The equivalence thought

These theories look different.

But a common thought is that the differences are merely notational, or at least merely matters of formulation.

> *Several versions of causal decision theory are on the market... But also I shall suggest that we causal decision theorists share one common idea, and differ mainly on matters of emphasis and formulation. The situation is not the chaos of disparate approaches that it may seem. (Lewis, 1981, p. 5)*

# The equivalence thought

These theories look different.

But a common thought is that the differences are merely notational, or at least merely matters of formulation.

> *Several versions of causal decision theory are on the market... But also I shall suggest that we causal decision theorists share one common idea, and differ mainly on matters of emphasis and formulation. The situation is not the chaos of disparate approaches that it may seem. (Lewis, 1981, p. 5)*

Today's question: can the Stalnakerian and Skyrmsian formulations really be equivalent?

# The bridge needed for equivalence

To identify Stalnakerian CDT with Skyrmsian CDT, we need a bridge between:

$$p(O \mathbin{\Box\!\!\rightarrow} S_i)$$

and

$$\sum_w p(w) \cdot ch_{H_{w,t^-}}(S_i \mid O).$$

# The bridge needed for equivalence

To identify Stalnakerian CDT with Skyrmsian CDT, we need a bridge between:

$$p(O \mathrel{\Box\!\!\rightarrow} S_i)$$

and

$$\sum_w p(w) \cdot ch_{H_{w,t^-}}(S_i \mid O).$$

That bridge is precisely the kind of principle we are about to discuss:

**Skyrms-style bridge**

$$p(O \mathrel{\Box\!\!\rightarrow} S_i) = \sum_w p(w) \cdot ch_{H_{w,t^-}}(S_i \mid O).$$

## The bridge needed for equivalence

To identify Stalnakerian CDT with Skyrmsian CDT, we need a bridge between:

$$p(O \mathrel{\Box\!\!\rightarrow} S_i)$$

and

$$\sum_w p(w) \cdot ch_{H_{w,t^-}}(S_i \mid O).$$

That bridge is precisely the kind of principle we are about to discuss:

**Skyrms-style bridge**

$$p(O \mathrel{\Box\!\!\rightarrow} S_i) = \sum_w p(w) \cdot ch_{H_{w,t^-}}(S_i \mid O).$$

The worry is that this bridge does not vindicate CDT. It threatens to make CDT collapse into ordinary Savage-style expected utility theory.

# Triviality for counterfactuals

- We saw: the assignment of probabilities to indicatives is problematic.
- What about counterfactuals?

# Counterfactuals and probability

- We saw: the assignment of probabilities to indicatives is problematic.
- What about counterfactuals?

Two problems (corresponding to two different clusters of literature).

i. Assigning intuitive probabilities to counterfactuals.

(1) If I tossed the coin, it would land tails.

The (potential) problem: we really need to make sure we use the right semantics! (E.g., Lewis semantics will make a lot of counterfactuals false.)

- We saw: the assignment of probabilities to indicatives is problematic.
- What about counterfactuals?

Two problems (corresponding to two different clusters of literature).

i. Assigning intuitive probabilities to counterfactuals.

(1) If I tossed the coin, it would land tails.

The (potential) problem: we really need to make sure we use the right semantics! (E.g., Lewis semantics will make a lot of counterfactuals false.)

ii. Triviality rearing its ugly head.

# Williams on counterfactual triviality

J.R.G. Williams (2012) presents a triviality result for conditional chances.

# Williams on counterfactual triviality

J.R.G. Williams (2012) presents a triviality result for conditional chances.

What do we mean by chance?

- Roughly: a prominent ‘objective’ interpretation of probability, which is supposed to figure in laws of nature.

# Williams on counterfactual triviality

J.R.G. Williams (2012) presents a triviality result for conditional chances.

What do we mean by chance?

- Roughly: a prominent ‘objective’ interpretation of probability, which is supposed to figure in laws of nature.
- Work in epistemology and metaphysics focuses on capturing the link between chance and credence. Williams’ proof exploits this bridge.

Following Meacham (2010), we take a chance function to have two arguments.

- The object argument is the proposition that is assigned a chance.
- The grounding argument is a proposition that specifies the information on the basis of which a probability value is assigned to the object argument.

# The chance function

Following Meacham (2010), we take a chance function to have two arguments.

- The object argument is the proposition that is assigned a chance.
- The grounding argument is a proposition that specifies the information on the basis of which a probability value is assigned to the object argument.

Simplifying, the grounding argument is a complete history of the world up to a time $t$ (a proposition specifying all matters of particular fact in $w$ up to $t$, $H_{w,t}$).

# The chance function

Following Meacham (2010), we take a chance function to have two arguments.

- The object argument is the proposition that is assigned a chance.
- The grounding argument is a proposition that specifies the information on the basis of which a probability value is assigned to the object argument.

Simplifying, the grounding argument is a complete history of the world up to a time $t$ (a proposition specifying all matters of particular fact in $w$ up to $t$, $H_{w,t}$).

$$ch_{H_{w,t}}(\mathrm{A})$$

The overall strategy.

i. Start from a counterfactual counterpart of the Thesis (i.e. a principle about credence).
ii. Derive a Thesis-like principle about chance.
iii. Motivate a closure assumption about chance.
iv. Run a Lewis-like triviality result using (i) and (ii).

Skyrms' Thesis.

Let $ch_{H_{w,t}}$ be the chance function at $w$ and $t$. For all A, C, and for all rational credence functions $cr_{E_t}$ such that $E$ is the subject's total evidence at $t$:

$$cr_{E_t}(\mathrm{A} \mathbin{\Box\!\!\rightarrow} \mathrm{C}) = \sum_{w \in W} cr_E(w) \times ch_{H_{w,t-}}(\mathrm{C} \mid \mathrm{A})$$

- Rough gloss: your (counterfactual) suppositional credence at $t$ in C, given A, should equal to your expectation of the conditional chance of $C$, given $A$, at a time that obtained 'just before' $t$.
- Your expectation is determined by your credences in candidates for the chance function in different worlds.

# Skyrms' Thesis

Example: the die. I have a fair die. I consider tossing it, but I don't.

Question: what is your credence in (2)?

(2) If I had tossed the die, it would have landed even.

Example: the die. I have a fair die. I consider tossing it, but I don't.

Question: what is your credence in (2)?

(2) If I had tossed the die, it would have landed even.

- Intuitive verdict: 1/2

Example: the die. I have a fair die. I consider tossing it, but I don't.

Question: what is your credence in (2)?

(2) If I had tossed the die, it would have landed even.

- Intuitive verdict: 1/2
- ST prediction: 1/2

# Skyrms' Thesis

Example: the die. I have a fair die. I consider tossing it, but I don't.

Question: what is your credence in (2)?

(2) If I had tossed the die, it would have landed even.

- Intuitive verdict: 1/2
- ST prediction: 1/2

Example: the coin. I have a coin that that is either double-tailed or double-headed, but you don't know which (and you think either option is .5 likely). I consider flipping it, but I don't.

Question: what is your credence in (3)?

(3) If I had flipped the coin, it would have landed heads.

Example: the die. I have a fair die. I consider tossing it, but I don't.

Question: what is your credence in (2)?

(2) If I had tossed the die, it would have landed even.

- Intuitive verdict: 1/2
- ST prediction: 1/2

Example: the coin. I have a coin that that is either double-tailed or double-headed, but you don't know which (and you think either option is .5 likely). I consider flipping it, but I don't.

Question: what is your credence in (3)?

(3) If I had flipped the coin, it would have landed heads.

- Intuitive verdict: 1/2

## Skyrms' Thesis

Example: the die. I have a fair die. I consider tossing it, but I don't.

Question: what is your credence in (2)?

(2) If I had tossed the die, it would have landed even.

- Intuitive verdict: 1/2
- ST prediction: 1/2

Example: the coin. I have a coin that that is either double-tailed or double-headed, but you don't know which (and you think either option is .5 likely). I consider flipping it, but I don't.

Question: what is your credence in (3)?

(3) If I had flipped the coin, it would have landed heads.

- Intuitive verdict: 1/2
- ST prediction: 1/2

Why believe Skyrms' Thesis?

# Williams on counterfactual triviality

Why believe Skyrms' Thesis?

Williams gives two reasons.

- Seems to get the results right in particular cases!
- It seems to give chance the right role to play in theorizing.

Why believe Skyrms' Thesis?

Williams gives two reasons.

- Seems to get the results right in particular cases!
- It seems to give chance the right role to play in theorizing.

Notice: Skyrms' Thesis is not the only principle cashing out subjunctive suppositional credences—more on this in a bit!

A bridge principle between credence and chance:

Principal Principle. (Lewis 1980)
$Pr(A|Ch(A) = x) = x$, assuming that $Pr(Ch(A) = x) > 0$

- This says: conditional on the chance of A being $x$, your credence in A should be $x$.
- Intuitively: you should take the chance function to be an expert on the probabilities to assign to propositions.
- Lewis adds a further caveat about 'inadmissible information' that we will ignore.

A bridge principle between credence and chance:

> Principal Principle. (Lewis 1980)
> $Pr(A|Ch(A) = x) = x$, assuming that $Pr(Ch(A) = x) > 0$

- This says: conditional on the chance of A being $x$, your credence in A should be $x$.
- Intuitively: you should take the chance function to be an expert on the probabilities to assign to propositions.
- Lewis adds a further caveat about 'inadmissible information' that we will ignore.

Example: suppose you're certain that the chance of Spain winning the WC is .3. Then your credence that Spain wins should be .3.

Assume a subject $s$ who is ideally informed about chances.

Assume a subject $s$ who is ideally informed about chances.

Fact: given the Principal Principle, Skyrms' Thesis is equivalent to:

Chancy Equation (CE). $ch_{H_{w,t}}(A \mathrel{\Box\!\!\rightarrow} C) = ch_{H_{w,t^-}}(C \mid A)$

# Williams on counterfactual triviality

Assume a subject $s$ who is ideally informed about chances.

Fact: given the Principal Principle, Skyrms' Thesis is equivalent to:

Chancy Equation (CE). $ch_{H_{w,t}}(A \mathrel{\Box\!\!\rightarrow} C) = ch_{H_{w,t^-}}(C \mid A)$

- What matters to use is one direction of the equivalence: Skyrms' Thesis entails the Chancy Equation (given PP).

Assume a subject $s$ who is ideally informed about chances.

Fact: given the Principal Principle, Skyrms' Thesis is equivalent to:

Chancy Equation (CE). $ch_{H_{w,t}}(A \mathrel{\Box\!\!\rightarrow} C) = ch_{H_{w,t^-}}(C \mid A)$

- What matters to use is one direction of the equivalence: Skyrms' Thesis entails the Chancy Equation (given PP).
- Here is why, intuitively: consider a subject has no uncertainty over the chances; the Chancy Equation immediately follows.

## Williams on counterfactual triviality

Assume a subject $s$ who is ideally informed about chances.

Fact: given the Principal Principle, Skyrms' Thesis is equivalent to:

Chancy Equation (CE). $ch_{H_{w,t}}(A \mathrel{\Box\!\!\rightarrow} C) = ch_{H_{w,t^-}}(C \mid A)$

- What matters to use is one direction of the equivalence: Skyrms' Thesis entails the Chancy Equation (given PP).
- Here is why, intuitively: consider a subject has no uncertainty over the chances; the Chancy Equation immediately follows.

Does this remind you of something?

One more principle: closure.

**Closure**

If $ch$ is the chance distribution at some world $w$ at time $t$, $\mathbf{A}$ any proposition (such that $ch(\mathbf{A}) > 0$), and $ch_0$ is the probability assignment that results from conditionalizing $ch$ on $\mathbf{A}$ (i.e. $ch_0(\cdot) = ch(\cdot\,|\mathbf{A})$), then there's a world $w_0$ and time $t_0$ whose chances at $t_0$ are described by $ch_0$.

One more principle: closure.

**Closure**

If $ch$ is the chance distribution at some world $w$ at time $t$, $\mathbf{A}$ any proposition (such that $ch(\mathbf{A}) > 0$), and $ch_0$ is the probability assignment that results from conditionalizing $ch$ on $\mathbf{A}$ (i.e. $ch_0(\cdot) = ch(\cdot \mid \mathbf{A})$), then there's a world $w_0$ and time $t_0$ whose chances at $t_0$ are described by $ch_0$.

One more principle: closure.

**Closure**

If $ch$ is the chance distribution at some world $w$ at time $t$, $\mathbf{A}$ any proposition (such that $ch(\mathbf{A}) > 0$), and $ch_0$ is the probability assignment that results from conditionalizing $ch$ on $\mathbf{A}$ (i.e. $ch_0(\cdot) = ch(\cdot\,|\mathbf{A})$), then there's a world $w_0$ and time $t_0$ whose chances at $t_0$ are described by $ch_0$.

This says, roughly: if you have a chance function, and you conditionalize on any propositions, you still have a chance function (of some world, at some time)!

One more principle: closure.

Closure

If $ch$ is the chance distribution at some world $w$ at time $t$, **A** any proposition (such that $ch(\mathbf{A}) > 0$), and $ch_0$ is the probability assignment that results from conditionalizing $ch$ on **A** (i.e. $ch_0(\cdot) = ch(\cdot\,|\mathbf{A})$), then there's a world $w_0$ and time $t_0$ whose chances at $t_0$ are described by $ch_0$.

This says, roughly: if you have a chance function, and you conditionalize on any propositions, you still have a chance function (of some world, at some time)!

Pet peeves:

- I suspect that closure is more demanding than Williams thinks!

One more principle: closure.

**Closure**

If $ch$ is the chance distribution at some world $w$ at time $t$, $\mathbf{A}$ any proposition (such that $ch(\mathbf{A}) > 0$), and $ch_0$ is the probability assignment that results from conditionalizing $ch$ on $\mathbf{A}$ (i.e. $ch_0(\cdot) = ch(\cdot\,|\mathbf{A})$), then there's a world $w_0$ and time $t_0$ whose chances at $t_0$ are described by $ch_0$.

This says, roughly: if you have a chance function, and you conditionalize on any propositions, you still have a chance function (of some world, at some time)!

Pet peeves:

- I suspect that closure is more demanding than Williams thinks!
- By conditionalizing on some propositions, you might land on a probability function that is not the chance function of any world.

## Williams on counterfactual triviality

One more principle: closure.

**Closure**

If $ch$ is the chance distribution at some world $w$ at time $t$, **A** any proposition (such that $ch(\mathbf{A}) > 0$), and $ch_0$ is the probability assignment that results from conditionalizing $ch$ on **A** (i.e. $ch_0(\cdot) = ch(\cdot\,|\mathbf{A})$), then there's a world $w_0$ and time $t_0$ whose chances at $t_0$ are described by $ch_0$.

This says, roughly: if you have a chance function, and you conditionalize on any propositions, you still have a chance function (of some world, at some time)!

Pet peeves:

- I suspect that closure is more demanding than Williams thinks!
- By conditionalizing on some propositions, you might land on a probability function that is not the chance function of any world.
- But I also think in the end these problems are not fatal.

Given all this, we can run a proof that resembles a lot the proof we gave yesterday.

## Williams on counterfactual triviality

Given all this, we can run a proof that resembles a lot the proof we gave yesterday.

i. $ch_{H_{w,t}}(A \mathrel{\Box\!\!\rightarrow} C) =$

Given all this, we can run a proof that resembles a lot the proof we gave yesterday.

i. $ch_{H_{w,t}}(\mathrm{A} \mathbin{\Box\!\!\rightarrow} \mathrm{C}) =$

ii. $ch_{H_{w,t}}(\mathrm{A} \mathbin{\Box\!\!\rightarrow} \mathrm{C} \mid \mathrm{C}) \times ch_{H_{w,t}}(\mathrm{C}) + ch_{H_{w,t}}(\mathrm{A} \mathbin{\Box\!\!\rightarrow} \mathrm{C} \mid \neg\mathrm{C}) \times ch_{H_{w,t}}(\neg\mathrm{C}) =$

Given all this, we can run a proof that resembles a lot the proof we gave yesterday.

i. $ch_{H_{w,t}}(\mathsf{A} \mathrel{\Box\!\!\rightarrow} \mathsf{C}) =$

ii. $ch_{H_{w,t}}(\mathsf{A} \mathrel{\Box\!\!\rightarrow} \mathsf{C} \mid \mathsf{C}) \times ch_{H_{w,t}}(\mathsf{C}) + ch_{H_{w,t}}(\mathsf{A} \mathrel{\Box\!\!\rightarrow} \mathsf{C} \mid \neg\mathsf{C}) \times ch_{H_{w,t}}(\neg\mathsf{C}) =$

iii. $ch_{H_{w,t}+\mathsf{C}}(\mathsf{A} \mathrel{\Box\!\!\rightarrow} \mathsf{C}) \times ch_{H_{w,t}}(\mathsf{C}) + ch_{H_{w,t}+\bar{\mathsf{C}}}(\mathsf{A} \mathrel{\Box\!\!\rightarrow} \mathsf{C}) \times ch_{H_{w,t}}(\neg\mathsf{C}) =$

Given all this, we can run a proof that resembles a lot the proof we gave yesterday.

i. $ch_{H_{w,t}}(\mathsf{A} \mathrel{\Box\!\!\rightarrow} \mathsf{C}) =$

ii. $ch_{H_{w,t}}(\mathsf{A} \mathrel{\Box\!\!\rightarrow} \mathsf{C} \mid \mathsf{C}) \times ch_{H_{w,t}}(\mathsf{C}) + ch_{H_{w,t}}(\mathsf{A} \mathrel{\Box\!\!\rightarrow} \mathsf{C} \mid \neg\mathsf{C}) \times ch_{H_{w,t}}(\neg\mathsf{C}) =$

iii. $ch_{H_{w,t}+\mathsf{c}}(\mathsf{A} \mathrel{\Box\!\!\rightarrow} \mathsf{C}) \times ch_{H_{w,t}}(\mathsf{C}) + ch_{H_{w,t}+\bar{\mathsf{c}}}(\mathsf{A} \mathrel{\Box\!\!\rightarrow} \mathsf{C}) \times ch_{H_{w,t}}(\neg\mathsf{C}) =$

iv. $ch_{H_{w,t^-}+\mathsf{c}}(\mathsf{C} \mid \mathsf{A}) \times ch_{H_{w,t}}(\mathsf{C}) + ch_{H_{w,t^-}+\bar{\mathsf{c}}}(\mathsf{C} \mid \mathsf{A}) \times ch_{H_{w,t}}(\neg\mathsf{C}) =$

Given all this, we can run a proof that resembles a lot the proof we gave yesterday.

i. $ch_{H_{w,t}}(A \mathrel{\Box\!\!\rightarrow} C) =$

ii. $ch_{H_{w,t}}(A \mathrel{\Box\!\!\rightarrow} C \mid C) \times ch_{H_{w,t}}(C) + ch_{H_{w,t}}(A \mathrel{\Box\!\!\rightarrow} C \mid \neg C) \times ch_{H_{w,t}}(\neg C) =$

iii. $ch_{H_{w,t}+\mathsf{C}}(A \mathrel{\Box\!\!\rightarrow} C) \times ch_{H_{w,t}}(C) + ch_{H_{w,t}+\bar{\mathsf{C}}}(A \mathrel{\Box\!\!\rightarrow} C) \times ch_{H_{w,t}}(\neg C) =$

iv. $ch_{H_{w,t^-}+\mathsf{C}}(C \mid A) \times ch_{H_{w,t}}(C) + ch_{H_{w,t^-}+\bar{\mathsf{C}}}(C \mid A) \times ch_{H_{w,t}}(\neg C) =$

v. $1 \times ch_{H_{w,t}}(C) + 0 \times ch_{H_{w,t}}(\neg C) =$

vi. $ch_{H_{w,t}}(C)$

# Williams on counterfactual triviality

i. $ch_{H_{w,t}}(A \mathrel{\Box\!\!\rightarrow} C) =$

ii. $ch_{H_{w,t}}(A \mathrel{\Box\!\!\rightarrow} C \mid C) \times ch_{H_{w,t}}(C) + ch_{H_{w,t}}(A \mathrel{\Box\!\!\rightarrow} C \mid \neg C) \times ch_{H_{w,t}}(\neg C)$

## Williams on counterfactual triviality

i. $ch_{H_{w,t}}(A \mathrel{\Box\!\!\rightarrow} C) =$

ii. $ch_{H_{w,t}}(A \mathrel{\Box\!\!\rightarrow} C \mid C) \times ch_{H_{w,t}}(C) + ch_{H_{w,t}}(A \mathrel{\Box\!\!\rightarrow} C \mid \neg C) \times ch_{H_{w,t}}(\neg C)$

- This step just uses the so-called Law of Total Probability.
- Essentially, we just use math and the definition of conditional probability.

# Williams on counterfactual triviality

ii. $ch_{H_{w,t}}(\mathrm{A} \mathrel{\Box\!\!\rightarrow} \mathrm{C} \mid \mathrm{C}) \times ch_{H_{w,t}}(\mathrm{C}) + ch_{H_{w,t}}(\mathrm{A} \mathrel{\Box\!\!\rightarrow} \mathrm{C} \mid \neg\mathrm{C}) \times ch_{H_{w,t}}(\neg\mathrm{C}) =$

iii. $ch_{H_{w,t}+\mathrm{C}}(\mathrm{A} \mathrel{\Box\!\!\rightarrow} \mathrm{C}) \times ch_{H_{w,t}}(\mathrm{C}) + ch_{H_{w,t}+\bar{\mathrm{C}}}(\mathrm{A} \mathrel{\Box\!\!\rightarrow} \mathrm{C}) \times ch_{H_{w,t}}(\neg\mathrm{C}) =$

# Williams on counterfactual triviality

ii. $ch_{H_{w,t}}(\mathrm{A} \mathrel{\Box\!\!\rightarrow} \mathrm{C} \mid \mathrm{C}) \times ch_{H_{w,t}}(\mathrm{C}) + ch_{H_{w,t}}(\mathrm{A} \mathrel{\Box\!\!\rightarrow} \mathrm{C} \mid \neg\mathrm{C}) \times ch_{H_{w,t}}(\neg\mathrm{C}) =$

iii. $ch_{H_{w,t}+\mathrm{C}}(\mathrm{A} \mathrel{\Box\!\!\rightarrow} \mathrm{C}) \times ch_{H_{w,t}}(\mathrm{C}) + ch_{H_{w,t}+\bar{\mathrm{C}}}(\mathrm{A} \mathrel{\Box\!\!\rightarrow} \mathrm{C}) \times ch_{H_{w,t}}(\neg\mathrm{C}) =$

- This step is justified by the idea that chances 'evolve by conditionalization'.
- When we conditionalize a chance function on a proposition, that is the same thing as conjoing that proposition to the grounding argument.

# Williams on counterfactual triviality

iii. $ch_{H_{w,t}+\mathsf{C}}(A \mathrel{\Box\!\!\rightarrow} C) \times ch_{H_{w,t}}(C) + ch_{H_{w,t}+\bar{\mathsf{C}}}(A \mathrel{\Box\!\!\rightarrow} C) \times ch_{H_{w,t}}(\neg C) =$

iv. $ch_{H_{w,t^-}+\mathsf{C}}(C \mid A) \times ch_{H_{w,t}}(C) + ch_{H_{w,t^-}+\bar{\mathsf{C}}}(C \mid A) \times ch_{H_{w,t}}(\neg C) =$

# Williams on counterfactual triviality

iii. $ch_{H_{w,t}+\mathsf{C}}(\mathsf{A} \mathrel{\Box\!\!\rightarrow} \mathsf{C}) \times ch_{H_{w,t}}(\mathsf{C}) + ch_{H_{w,t}+\bar{\mathsf{C}}}(\mathsf{A} \mathrel{\Box\!\!\rightarrow} \mathsf{C}) \times ch_{H_{w,t}}(\neg\mathsf{C}) =$

iv. $ch_{H_{w,t^-}+\mathsf{C}}(\mathsf{C} \mid \mathsf{A}) \times ch_{H_{w,t}}(\mathsf{C}) + ch_{H_{w,t^-}+\bar{\mathsf{C}}}(\mathsf{C} \mid \mathsf{A}) \times ch_{H_{w,t}}(\neg\mathsf{C}) =$

- This is where we apply the Chancy Equation.
- Notice: we replace chances of conditionals with conditional chances.

iv. $ch_{H_{w,t^-}+C}(C \mid A) \times ch_{H_{w,t}}(C) + ch_{H_{w,t^-}+\bar{C}}(C \mid A) \times ch_{H_{w,t}}(\neg C) =$

v. $1 \times ch_{H_{w,t}}(C) + 0 \times ch_{H_{w,t}}(\neg C)$

## Williams on counterfactual triviality

iv. $ch_{H_{w,t^-}+\mathsf{C}}(\mathsf{C} \mid \mathsf{A}) \times ch_{H_{w,t}}(\mathsf{C}) + ch_{H_{w,t^-}+\bar{\mathsf{C}}}(\mathsf{C} \mid \mathsf{A}) \times ch_{H_{w,t}}(\neg \mathsf{C}) =$

v. $1 \times ch_{H_{w,t}}(\mathsf{C}) + 0 \times ch_{H_{w,t}}(\neg \mathsf{C})$

- If history includes the information that C, the chance of C is one.
- Similarly, *mutatis mutandis*, for ¬C.

How to respond?

# Williams on counterfactual triviality

How to respond?

(i) Accept the result (and give up on the equivalence of different decision theories).

How to respond?

(i) Accept the result (and give up on the equivalence of different decision theories).

(ii) Block some principle in the proof (Closure? Conditionalization?).

How to respond?

(i) Accept the result (and give up on the equivalence of different decision theories).

(ii) Block some principle in the proof (Closure? Conditionalization?).

(iii) Reject Skyrms' Thesis, and possibly look for an alternative.

A way of going the way of (iii): Santorio 2025.

There I propose to replace the Chancy Equation with the following:

**Counterfactual Chance Thesis**

$ch_{H_{w,t}}(\mathsf{A} \mathrel{\Box\!\!\rightarrow}_s \mathsf{C}) = ch_{s(\mathsf{A},H_{w,t})}(\mathsf{C})$

A way of going the way of (iii): Santorio 2025.

There I propose to replace the Chancy Equation with the following:

**Counterfactual Chance Thesis**

$ch_{H_{w,t}}(\mathsf{A} \mathrel{\Box\!\!\rightarrow_s} \mathsf{C}) = ch_{s(\mathsf{A}, H_{w,t})}(\mathsf{C})$

- This says: the chance that $\mathsf{A} \mathrel{\Box\!\!\rightarrow_s} \mathsf{C}$ equals the chance of C, in the closest world history where A is the case.
- Crucially, we use the same selection function for the counterfactual and for the grounding argument of the chance function.
- In a slogan: chances of counterfactuals are counterfactual chances.

## The Counterfactual Chance Thesis

How does this block triviality?

How does this block triviality?

In Williams' proof, the triviality-generating step is the equality in (4):

(4) $ch_{T_w H_{w,t}+\mathsf{C}}(\mathsf{A} \,\Box\!\!\rightarrow \mathsf{C}) = ch_{T_w H_{w,t}+\mathsf{C}}(\mathsf{C} \mid \mathsf{A})$

(4) The right-hand side of (4) goes to 1 for mathematical reasons.

How does this block triviality?

In Williams' proof, the triviality-generating step is the equality in (4):

(4) $ch_{T_w H_{w,t} + \mathsf{C}}(\mathsf{A} \Box\!\!\rightarrow \mathsf{C}) = ch_{T_w H_{w,t} + \mathsf{C}}(\mathsf{C} \mid \mathsf{A})$

(4) The right-hand side of (4) goes to 1 for mathematical reasons.

If we switch to CCT, (4) is replaced by:

(4) $ch_{T_w H_{w,t} + \mathsf{C}}(\mathsf{A} \Box\!\!\rightarrow \mathsf{C}) = ch_{T_w s(\mathsf{A}, H_{w,t} + \mathsf{C})}(\mathsf{C})$

- Crucially, the right-hand side of 4 need *not* be identical to 1.
- The reason is that, even if we start from a history that vindicates C, the selection function can take us to a history where C is not true/not settled.

# The Counterfactual Chance Thesis

For illustration, consider this scenario:

Two coins. Sarah is going to flip one of two coins A and B, picked at random. Coin A is fair. Coin B has a 100% bias towards heads.

- Let 'A-Flip' stand for the proposition that coin A is flipped, and 'Heads' be the proposition that whichever coin is flipped lands heads.
- Now, consider the chance of the counterfactual A-Flip $\Box\!\!\rightarrow$ Heads, under the hypothesis that the flipped coin lands heads.

(5) $\quad ch_{H_{w,t}+\textbf{Heads}}(\text{A-Flip} \Box\!\!\rightarrow \text{Heads})$

- Via the Chancy Equation, this is 1.

For illustration, consider this scenario:

Two coins. Sarah is going to flip one of two coins A and B, picked at random. Coin A is fair. Coin B has a 100% bias towards heads.

- Let 'A-Flip' stand for the proposition that coin A is flipped, and 'Heads' be the proposition that whichever coin is flipped lands heads.
- Now, consider the chance of the counterfactual A-Flip $\Box\!\!\rightarrow$ Heads, under the hypothesis that the flipped coin lands heads.

(5) $ch_{H_{w,t}+\textbf{Heads}}(\text{A-Flip} \mathrel{\Box\!\!\rightarrow} \text{Heads})$

- Via the Chancy Equation, this is 1.
- Intuitively, . . .

For illustration, consider this scenario:

Two coins. Sarah is going to flip one of two coins A and B, picked at random. Coin A is fair. Coin B has a 100% bias towards heads.

- Let 'A-Flip' stand for the proposition that coin A is flipped, and 'Heads' be the proposition that whichever coin is flipped lands heads.
- Now, consider the chance of the counterfactual A-Flip $\Box\!\!\rightarrow$ Heads, under the hypothesis that the flipped coin lands heads.

(5) $ch_{H_{w,t}+\textbf{Heads}}(\text{A-Flip} \Box\!\!\rightarrow \text{Heads})$

- Via the Chancy Equation, this is 1.
- Intuitively, . . . it is less than 1.

# The Counterfactual Chance Thesis

For illustration, consider this scenario:

Two coins. Sarah is going to flip one of two coins A and B, picked at random. Coin A is fair. Coin B has a 100% bias towards heads.

- Let 'A-Flip' stand for the proposition that coin A is flipped, and 'Heads' be the proposition that whichever coin is flipped lands heads.
- Now, consider the chance of the counterfactual A-Flip $\Box\!\!\rightarrow$ Heads, under the hypothesis that the flipped coin lands heads.

(5) $ch_{H_{w,t}+\textbf{Heads}}(\text{A-Flip} \Box\!\!\rightarrow \text{Heads})$

- Via the Chancy Equation, this is 1.
- Intuitively, ... it is less than 1.
- The CCT allows it to be less than 1.

(5) $ch_{H_{w,t}+\textbf{Heads}}(\text{A-Flip} \mathrel{\Box\!\!\rightarrow} \text{Heads})$

Given the CCT, (5) equals:

(6) $ch_{s(\textbf{A-Flip},H_{w,t}+\textbf{Heads})}(\text{Heads})$

- Crucially, $s(\text{A-Flip}, H_{w,t} + \text{Heads})$ need not be a scenario that makes heads true.
- In fact, on a plausible choice of selection function, some closest worlds where A is flipped will lead to a tails outcome.
- Hence the chance of heads will be less than 1, hence the chance of A-Flip $\Box\!\!\rightarrow$ Heads, under the hypothesis that the flipped coin lands heads, is less than 1.

## The Counterfactual Chance Thesis

So: we have a recipe to block triviality!

But there is work to do:

- No tenability result for the CCT (yet).
- No idea how it interacts with decision theory.

# Causal decision theory and triviality

## Three versions of CDT

Recall the three versions of CDT we looked at:

# Three versions of CDT

Recall the three versions of CDT we looked at:

- Stalnakerian CDT

$$U_{\mathrm{ST}}(O) = \sum_i p(O \mathrel{\Box\!\!\rightarrow} S_i) \cdot u(O \cap S_i)$$

Recall the three versions of CDT we looked at:

- Stalnakerian CDT

$$U_{\mathrm{ST}}(O) = \sum_i p(O \boxright S_i) \cdot u(O \cap S_i)$$

- Lewisian CDT

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

where the $K_i$ are dependency hypotheses.

Recall the three versions of CDT we looked at:

- Stalnakerian CDT

$$U_{\mathrm{ST}}(O) = \sum_i p(O \mathrel{\Box\!\!\rightarrow} S_i) \cdot u(O \cap S_i)$$

- Lewisian CDT

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

where the $K_i$ are dependency hypotheses.

- Skyrmsian CDT

$$U_{\mathrm{SK}}(O) = \sum_i \sum_w p(w) \cdot ch_{H_{w,t^-}}(S_i \mid O) \cdot u(O \cap S_i)$$

# The equivalence thought

Once again, these theories look different.

# The equivalence thought

Once again, these theories look different.

But they are often claimed to be equivalent:

Once again, these theories look different.

But they are often claimed to be equivalent:

> *Several versions of causal decision theory are on the market... But also I shall suggest that we causal decision theorists share one common idea, and differ mainly on matters of emphasis and formulation. The situation is not the chaos of disparate approaches that it may seem. (Lewis, 1981, p. 5)*

# Equivalence?

Here's an even more explicit statement of the equivalence claim, from Joyce (1999):

> *From a purely decision-theoretic perspective, the similarities among the "causally relevant factor," "conditional chance," and "counterfactual dependence" versions of [CDT] are more striking than their differences. A person's judgment about which of the three interpretations is correct will depend on her views about the metaphysics and epistemology of causation rather than her intuitions about how to make choices.... Skyrms and Harper have expressed the same sentiment, observing, "**It can be argued that the various forms of causal decision theory are equivalent**—that an adequate version of any one of [them] will be interdefinable with adequate versions of the others." I think this is basically right; **there is no great difference between any of the approaches we have discussed**. (p. 171)*

# From counterfactual triviality to CDT

The triviality result for Skyrms's thesis, however, concerns exactly the bridge that the equivalence between these versions of CDT needs.

# From counterfactual triviality to CDT

The triviality result for Skyrms's thesis, however, concerns exactly the bridge that the equivalence between these versions of CDT needs.

Suppressing time indices, the Chancy Equation says:

$$ch(A \mathrel{\Box\!\!\rightarrow} C) = ch(C \mid A).$$

## From counterfactual triviality to CDT

The triviality result for Skyrms's thesis, however, concerns exactly the bridge that the equivalence between these versions of CDT needs.

Suppressing time indices, the Chancy Equation says:

$$ch(A \mathrel{\Box\!\!\rightarrow} C) = ch(C \mid A).$$

But the triviality result gives:

$$ch(A \mathrel{\Box\!\!\rightarrow} C) = ch(C).$$

The triviality result for Skyrms's thesis, however, concerns exactly the bridge that the equivalence between these versions of CDT needs.

Suppressing time indices, the Chancy Equation says:

$$ch(A \mathbin{\Box\!\!\rightarrow} C) = ch(C \mid A).$$

But the triviality result gives:

$$ch(A \mathbin{\Box\!\!\rightarrow} C) = ch(C).$$

Putting these together:

**Triviality**

$$ch(C \mid A) = ch(C).$$

## From counterfactual triviality to CDT

The triviality result for Skyrms's thesis, however, concerns exactly the bridge that the equivalence between these versions of CDT needs.

Suppressing time indices, the Chancy Equation says:

$$ch(A \mathrel{\Box\!\!\rightarrow} C) = ch(C \mid A).$$

But the triviality result gives:

$$ch(A \mathrel{\Box\!\!\rightarrow} C) = ch(C).$$

Putting these together:

**Triviality**

$$ch(C \mid A) = ch(C).$$

So, if the result holds, antecedents make no difference to the chances of their consequents.

## Apply this to options and states

Now let the antecedent be an option $O$, and let the consequent be a state $S_i$.

## Apply this to options and states

Now let the antecedent be an option $O$, and let the consequent be a state $S_i$.

Then triviality says:

$$ch(S_i \mid O) = ch(S_i).$$

# Apply this to options and states

Now let the antecedent be an option $O$, and let the consequent be a state $S_i$.

Then triviality says:

$$ch(S_i \mid O) = ch(S_i).$$

But Skyrmsian CDT evaluates $O$ by conditional chances:

$$U_{\mathrm{SK}}(O) = \sum_i \sum_w p(w) \cdot ch_{H_{w,t^-}}(S_i \mid O) \cdot u(O \cap S_i).$$

# Apply this to options and states

Now let the antecedent be an option $O$, and let the consequent be a state $S_i$.

Then triviality says:

$$ch(S_i \mid O) = ch(S_i).$$

But Skyrmsian CDT evaluates $O$ by conditional chances:

$$U_{\mathrm{SK}}(O) = \sum_i \sum_w p(w) \cdot ch_{H_{w,t^-}}(S_i \mid O) \cdot u(O \cap S_i).$$

By triviality, this becomes:

$$U_{\mathrm{SK}}(O) = \sum_i \sum_w p(w) \cdot ch_{H_{w,t^-}}(S_i) \cdot u(O \cap S_i).$$

If your credences defer to the relevant chances, then:

$$\sum_{w} p(w) \cdot ch_{H_{w,t-}}(S_i) = p(S_i).$$

# Collapse into Savage

If your credences defer to the relevant chances, then:

$$\sum_{w} p(w) \cdot ch_{H_{w,t-}}(S_i) = p(S_i).$$

So Skyrmsian CDT collapses into:

$$U_{\mathrm{SK}}(O) = \sum_{i} p(S_i) \cdot u(O \cap S_i).$$

# Collapse into Savage

If your credences defer to the relevant chances, then:

$$\sum_w p(w) \cdot ch_{H_{w,t-}}(S_i) = p(S_i).$$

So Skyrmsian CDT collapses into:

$$U_{\mathrm{SK}}(O) = \sum_i p(S_i) \cdot u(O \cap S_i).$$

But that is just Savage-style expected utility theory: attach probabilities to states independently of which option you choose.

# Collapse into Savage

If your credences defer to the relevant chances, then:

$$\sum_{w} p(w) \cdot ch_{H_{w,t-}}(S_i) = p(S_i).$$

So Skyrmsian CDT collapses into:

$$U_{\mathrm{SK}}(O) = \sum_{i} p(S_i) \cdot u(O \cap S_i).$$

But that is just Savage-style expected utility theory: attach probabilities to states independently of which option you choose.

**Upshot**

If the triviality result for Skyrms's thesis holds, then the bridge CDT'ers need does not vindicate CDT. It makes CDT collapse into Savage!

## The moral

The problem is not a new derivation.

## The moral

The problem is not a new derivation.

It is an application of the triviality result for Skyrms's thesis to decision theory.

## The moral

The problem is not a new derivation.

It is an application of the triviality result for Skyrms's thesis to decision theory.

If Stalnakerian CDT is equivalent to Skyrmsian CDT, and Skyrmsian CDT collapses into Savage, then Stalnakerian CDT collapses too.

# The moral

The problem is not a new derivation.

It is an application of the triviality result for Skyrms's thesis to decision theory.

If Stalnakerian CDT is equivalent to Skyrmsian CDT, and Skyrmsian CDT collapses into Savage, then Stalnakerian CDT collapses too.

So the equivalence claim is costly. Something has to give:

- the bridge from counterfactuals to conditional chances;
- the Principal Principle;
- closure for chance functions;
- or the claim that these CDT formulations are equivalent.

# Bonus: Desire-as-Belief

## Triviality in ethics?

We are going to close the course with a surprising application of the triviality results we've been discussing.

# Triviality in ethics?

We are going to close the course with a surprising application of the triviality results we've been discussing.

In particular, we'll look at a view known as *Desire-as-Belief*:

# Triviality in ethics?

We are going to close the course with a surprising application of the triviality results we've been discussing.

In particular, we'll look at a view known as *Desire-as-Belief*:

Desire-as-Belief: The degree to which you *desire* the truth of $A$ should be equal to your degree of belief that $A$'s truth would be good.

# Triviality in ethics?

We are going to close the course with a surprising application of the triviality results we've been discussing.

In particular, we'll look at a view known as *Desire-as-Belief*:

Desire-as-Belief: The degree to which you *desire* the truth of $A$ should be equal to your degree of belief that $A$'s truth would be good.

This view has played an important role in ethics (Parfit, 1984).

# Triviality in ethics?

We are going to close the course with a surprising application of the triviality results we've been discussing.

In particular, we'll look at a view known as *Desire-as-Belief*:

Desire-as-Belief: The degree to which you *desire* the truth of $A$ should be equal to your degree of belief that $A$'s truth would be good.

This view has played an important role in ethics (Parfit, 1984).

We are going to show, however, that in the context of decision theory, it faces stark challenges.

# Evidential decision theory, reconsidered

First, however, let us return to evidential decision theory.

## Evidential decision theory, reconsidered

First, however, let us return to evidential decision theory.

Recall our equation for that theory:

First, however, let us return to evidential decision theory.

Recall our equation for that theory:

**Evidential expected utility**

$$\mathrm{EEU}(O) = \sum_i p(S_i \mid O) \cdot u(O \cap S_i)$$

# Evidential decision theory, reconsidered

First, however, let us return to evidential decision theory.

Recall our equation for that theory:

**Evidential expected utility**

$$\mathrm{EEU}(O) = \sum_i p(S_i \mid O) \cdot u(O \cap S_i)$$

We have argued that evidential decision theory isn't the correct decision theory.

# Evidential decision theory, reconsidered

First, however, let us return to evidential decision theory.

Recall our equation for that theory:

**Evidential expected utility**

$$\mathrm{EEU}(O) = \sum_i p(S_i \mid O) \cdot u(O \cap S_i)$$

We have argued that evidential decision theory isn't the correct decision theory.

There is, however, a sense in which it's still useful...

To see this, consider this quote from Lewis (1996):

## Value vs choiceworthiness

To see this, consider this quote from Lewis (1996):

> *Suppose a certain action would serve as an effective means to your ends. Yet at the same time it would constitute evidence—evidence available to you in no other way that you are predestined inescapably to some dreadful misfortune. Should you perform that action?—Yes; your destiny is not a consideration,since that is outside your control. Do you desire to perform it?—No; you want good news, not bad. Since our topic here is not choiceworthiness but desire, and since the two diverge, we adopt an "evidential" conception of expected value, on which the value of the useful action that brings bad news is low. Choiceworthiness is governed by a different, "causal", conception of expected value.*

To see this, consider this quote from Lewis (1996):

> *Suppose a certain action would serve as an effective means to your ends. Yet at the same time it would constitute evidence—evidence available to you in no other way that you are predestined inescapably to some dreadful misfortune. Should you perform that action?—Yes; your destiny is not a consideration,since that is outside your control. Do you desire to perform it?—No; you want good news, not bad. Since our topic here is not choiceworthiness but desire, and since the two diverge, we adopt an "evidential" conception of expected value, on which the value of the useful action that brings bad news is low. Choiceworthiness is governed by a different, "causal", conception of expected value.*

Lewis is here contrasting evidential and causal decision theory.

To see this, consider this quote from Lewis (1996):

> *Suppose a certain action would serve as an effective means to your ends. Yet at the same time it would constitute evidence—evidence available to you in no other way that you are predestined inescapably to some dreadful misfortune. Should you perform that action?—Yes; your destiny is not a consideration,since that is outside your control. Do you desire to perform it?—No; you want good news, not bad. Since our topic here is not choiceworthiness but desire, and since the two diverge, we adopt an "evidential" conception of expected value, on which the value of the useful action that brings bad news is low. Choiceworthiness is governed by a different, "causal", conception of expected value.*

Lewis is here contrasting evidential and causal decision theory.

He is saying that EDT is not the right theory of how you should *choose*; but it may still be the right theory of rational subjective value.

# How to read the equation

To make this explicit, let us write $V(A)$ for the evidential expected utility of $A$, considered as the *rational subjective value* you should attach to $A$. In the language of worlds, we have:

$$V(A) = \sum_w p(w \mid A) \cdot u(w)$$

# How to read the equation

To make this explicit, let us write $V(A)$ for the evidential expected utility of $A$, considered as the *rational subjective value* you should attach to $A$. In the language of worlds, we have:

$$V(A) = \sum_{w} p(w \mid A) \cdot u(w)$$

We can think of it as telling you: how good things would look if you were to learn that $A$ is true.

# How to read the equation

To make this explicit, let us write $V(A)$ for the evidential expected utility of $A$, considered as the *rational subjective value* you should attach to $A$. In the language of worlds, we have:

$$V(A) = \sum_{w} p(w \mid A) \cdot u(w)$$

We can think of it as telling you: how good things would look if you were to learn that $A$ is true.

When $A$ is an available option, $O$, maximizing this quantity is just evidential decision theory.

Now let's turn to Desire-as-Belief (DAB).

# Binary goodness

Now let's turn to Desire-as-Belief (DAB).

To simplify, suppose worlds divide into two groups:

Now let's turn to Desire-as-Belief (DAB).

To simplify, suppose worlds divide into two groups:

| Kind of world | Proposition | Value |
|---|---|---|
| Good worlds | $G$ | 1 |
| Bad worlds | $\neg G$ | 0 |

# Binary goodness

Now let's turn to Desire-as-Belief (DAB).

To simplify, suppose worlds divide into two groups:

| Kind of world | Proposition | Value |
|---|---|---|
| Good worlds | $G$ | 1 |
| Bad worlds | $\neg G$ | 0 |

Then our equation for evidential expected utility yields:

$$\begin{aligned} V(A) &= p(G \mid A) \cdot V(G) + p(\neg G \mid A) \cdot V(\neg G) \\ &= p(G \mid A) \cdot 1 + p(\neg G \mid A) \cdot 0 \\ &= p(G \mid A). \end{aligned}$$

# Binary goodness

Now let's turn to Desire-as-Belief (DAB).

To simplify, suppose worlds divide into two groups:

| Kind of world | Proposition | Value |
|---|---|---|
| Good worlds | $G$ | 1 |
| Bad worlds | $\neg G$ | 0 |

Then our equation for evidential expected utility yields:

$$\begin{aligned} V(A) &= p(G \mid A) \cdot V(G) + p(\neg G \mid A) \cdot V(\neg G) \\ &= p(G \mid A) \cdot 1 + p(\neg G \mid A) \cdot 0 \\ &= p(G \mid A). \end{aligned}$$

So, in the binary case, the value of $A$ is the probability that things are good, conditional on $A$.

# Binary goodness

Now let's turn to Desire-as-Belief (DAB).

To simplify, suppose worlds divide into two groups:

| Kind of world | Proposition | Value |
|---|---|---|
| Good worlds | $G$ | 1 |
| Bad worlds | $\neg G$ | 0 |

Then our equation for evidential expected utility yields:

$$\begin{aligned} V(A) &= p(G \mid A) \cdot V(G) + p(\neg G \mid A) \cdot V(\neg G) \\ &= p(G \mid A) \cdot 1 + p(\neg G \mid A) \cdot 0 \\ &= p(G \mid A). \end{aligned}$$

So, in the binary case, the value of $A$ is the probability that things are good, conditional on $A$.

So far, so good...

# Desire-as-Belief

# The Desire-as-Belief Thesis

The Desire-as-Belief (DAB) Thesis is a claim about the relationship between your subjective values (utilities), on the one hand, and your beliefs about what's good, on the other.

# The Desire-as-Belief Thesis

The Desire-as-Belief (DAB) Thesis is a claim about the relationship between your subjective values (utilities), on the one hand, and your beliefs about what's good, on the other.

Let $A^\circ$ be the proposition expressed by: '$A$'s truth would be good.'

# The Desire-as-Belief Thesis

The Desire-as-Belief (DAB) Thesis is a claim about the relationship between your subjective values (utilities), on the one hand, and your beliefs about what's good, on the other.

Let $A^\circ$ be the proposition expressed by: '$A$'s truth would be good.'

Then Lewis's Desire-as-Belief thesis says (Lewis, 1988, 1996):

**DAB Thesis**

$$V(A) = p(A^\circ).$$

The Desire-as-Belief (DAB) Thesis is a claim about the relationship between your subjective values (utilities), on the one hand, and your beliefs about what's good, on the other.

Let $A^\circ$ be the proposition expressed by: '$A$'s truth would be good.'

Then Lewis's Desire-as-Belief thesis says (Lewis, 1988, 1996):

**DAB Thesis**

$$V(A) = p(A^\circ).$$

In words: Your degree of desire for $A$ should match your degree of belief that $A$'s truth would be *good*.

# Why DAB is tempting

DAB says that desire is constrained by belief about value.

# Why DAB is tempting

DAB says that desire is constrained by belief about value.

For realists about value, this is a natural thought:

# Why DAB is tempting

DAB says that desire is constrained by belief about value.

For realists about value, this is a natural thought:

- there are facts about what is good;

# Why DAB is tempting

DAB says that desire is constrained by belief about value.

For realists about value, this is a natural thought:

- there are facts about what is good;
- rational belief should track those facts;

## Why DAB is tempting

DAB says that desire is constrained by belief about value.

For realists about value, this is a natural thought:

- there are facts about what is good;
- rational belief should track those facts;
- rational desire should track one's beliefs about those facts.

## Why DAB is tempting

DAB says that desire is constrained by belief about value.

For realists about value, this is a natural thought:

- there are facts about what is good;
- rational belief should track those facts;
- rational desire should track one's beliefs about those facts.

So DAB can look like a value-theoretic analogue of the Principal Principle:

subjective attitudes should match beliefs about objective magnitudes.

# Lewis's ingredients

Lewis's argument uses three ingredients.

Lewis's argument uses three ingredients.

1. DAB: rational agents satisfy $V(A) = p(A^{\circ})$.

Lewis's argument uses three ingredients.

1. DAB: rational agents satisfy $V(A) = p(A^\circ)$.
2. Conditionalization: after learning $A$, your credence is

$$p_A(X) = p(X \mid A).$$

# Lewis's ingredients

Lewis's argument uses three ingredients.

1. DAB: rational agents satisfy $V(A) = p(A^\circ)$.
2. Conditionalization: after learning $A$, your credence is

$$p_A(X) = p(X \mid A).$$

3. Invariance: learning that $A$ does not change the value of $A$:

$$v_A(A) = V(A).$$

# Lewis's ingredients

Lewis's argument uses three ingredients.

1. DAB: rational agents satisfy $V(A) = p(A^\circ)$.
2. Conditionalization: after learning $A$, your credence is

$$p_A(X) = p(X \mid A).$$

3. Invariance: learning that $A$ does not change the value of $A$:

$$v_A(A) = V(A).$$

The question is: can all three hold generally?

No straightforwardly. Assume DAB holds in general (before and after learning). Then:

No straightforwardly. Assume DAB holds in general (before and after learning). Then:

$$\begin{aligned} p(A^\circ) &= V(A) && \text{DAB} \\ &= v_A(A) && \text{Invariance} \\ &= p_A(A^\circ) && \text{DAB again} \\ &= p(A^\circ \mid A) && \text{Conditionalization.} \end{aligned}$$

## Lewis's argument

No straightforwardly. Assume DAB holds in general (before and after learning). Then:

$$
\begin{aligned}
p(A^\circ) &= V(A) && \text{DAB} \\
&= v_A(A) && \text{Invariance} \\
&= p_A(A^\circ) && \text{DAB again} \\
&= p(A^\circ \mid A) && \text{Conditionalization.}
\end{aligned}
$$

So DAB seems to require:

**Independence**

$$p(A^\circ) = p(A^\circ \mid A).$$

No straightforwardly. Assume DAB holds in general (before and after learning). Then:

$$\begin{aligned} p(A^\circ) &= V(A) && \text{DAB} \\ &= v_A(A) && \text{Invariance} \\ &= p_A(A^\circ) && \text{DAB again} \\ &= p(A^\circ \mid A) && \text{Conditionalization.} \end{aligned}$$

So DAB seems to require:

**Independence**

$$p(A^\circ) = p(A^\circ \mid A).$$

But, in general, $A$ and $A^\circ$ need not be probabilistically independent.

## Why this is a problem

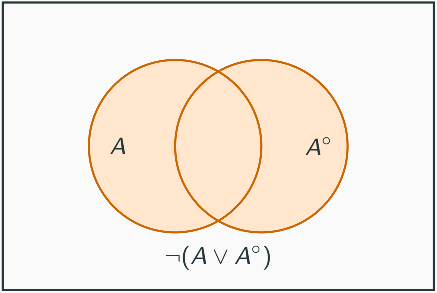
Suppose initially $A$ and $A^\circ$ are independent:

$$p(A^\circ) = p(A^\circ \mid A).$$

# Why this is a problem

Suppose initially $A$ and $A^\circ$ are independent:

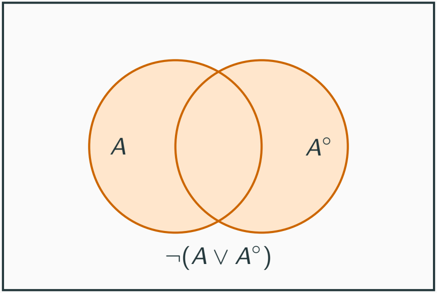
$$p(A^\circ) = p(A^\circ \mid A).$$



# Why this is a problem

Suppose initially $A$ and $A^\circ$ are independent:

$$p(A^\circ) = p(A^\circ \mid A).$$

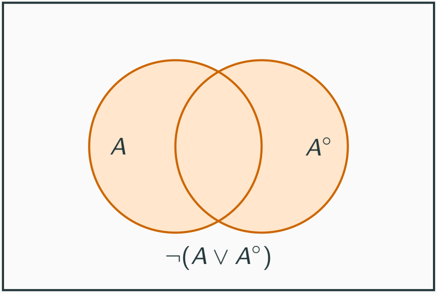



Now learn $A \vee A^\circ$ and update by Conditionalization.

# Why this is a problem

Suppose initially $A$ and $A^\circ$ are independent:

$$p(A^\circ) = p(A^\circ \mid A).$$



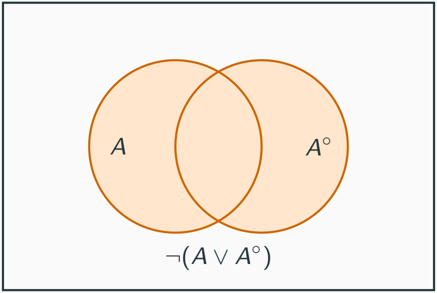

Now learn $A \vee A^\circ$ and update by Conditionalization.

- $p(A^\circ \mid A)$ stays fixed.

# Why this is a problem

Suppose initially $A$ and $A^\circ$ are independent:

$$p(A^\circ) = p(A^\circ \mid A).$$



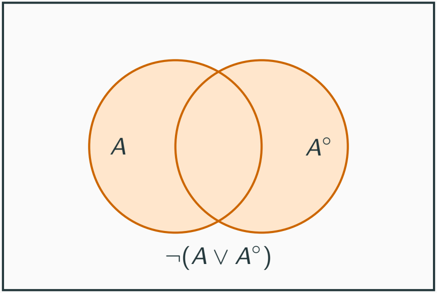

Now learn $A \vee A^\circ$ and update by Conditionalization.

- $p(A^\circ \mid A)$ stays fixed.
- $p(A^\circ)$ increases.

# Why this is a problem

Suppose initially $A$ and $A^\circ$ are independent:

$$p(A^\circ) = p(A^\circ \mid A).$$




Now learn $A \vee A^\circ$ and update by Conditionalization.

- $p(A^\circ \mid A)$ stays fixed.
- $p(A^\circ)$ increases.

So DAB can go from satisfied to unsatisfied.

# First moral

The problem is not just that DAB sometimes fails.

## First moral

The problem is not just that DAB sometimes fails.

The problem is that, if we combine DAB with Conditionalization and Invariance, then DAB imposes an extremely strong constraint:

$$p(A^\circ) = p(A^\circ \mid A)$$

for arbitrary $A$.

# First moral

The problem is not just that DAB sometimes fails.

The problem is that, if we combine DAB with Conditionalization and Invariance, then DAB imposes an extremely strong constraint:

$$p(A^{\circ}) = p(A^{\circ} \mid A)$$

for arbitrary $A$.

That is the beginning of a triviality result: the thesis can hold only in highly special cases.

# Triviality

# Stalnaker's Thesis

Yesterday, we discussed a famous thesis about indicatives.

Yesterday, we discussed a famous thesis about indicatives.

**Stalnaker's Thesis**

$$p(A \rightarrow B) = p(B \mid A).$$

# Stalnaker's Thesis

Yesterday, we discussed a famous thesis about indicatives.

**Stalnaker's Thesis**

$$p(A \rightarrow B) = p(B \mid A).$$

Your credence in *If A, then B* should match your conditional credence in $B$, given $A$ (Stalnaker, 1970; Lewis, 1976).

Yesterday, we discussed a famous thesis about indicatives.

**Stalnaker's Thesis**

$$p(A \to B) = p(B \mid A).$$

Your credence in *If A, then B* should match your conditional credence in $B$, given $A$ (Stalnaker, 1970; Lewis, 1976).

Example:

*If the die didn't land on* 1, *it landed on* 2.

Yesterday, we discussed a famous thesis about indicatives.

**Stalnaker's Thesis**

$$p(A \to B) = p(B \mid A).$$

Your credence in *If A, then B* should match your conditional credence in $B$, given $A$ (Stalnaker, 1970; Lewis, 1976).

Example:

*If the die didn't land on* 1*, it landed on* 2*.*

For a fair die, the natural credence is:

$$p(2 \mid \neg 1) = \frac{1}{5}.$$

We heard yesterday, however, that if Stalnaker's Thesis is tru—and given some other plausible assumptions—then

$$p(B \mid A) = p(B).$$

That looks quite a bit like the conclusion of our argument against the Desire-as-Belief Thesis.

# Why this is triviality

The conclusion says that $A$ is probabilistically irrelevant to $B$.

$$p(B \mid A) = p(B)$$

# Why this is triviality

The conclusion says that $A$ is probabilistically irrelevant to $B$.

$$p(B \mid A) = p(B)$$

But Stalnaker's Thesis was supposed to hold for many pairs of propositions $A, B$.

# Why this is triviality

The conclusion says that $A$ is probabilistically irrelevant to $B$.

$$p(B \mid A) = p(B)$$

But Stalnaker's Thesis was supposed to hold for many pairs of propositions $A, B$.

So, if the derivation is sound, then Stalnaker's Thesis can hold generally only in trivial probability spaces.

# Why this is triviality

The conclusion says that $A$ is probabilistically irrelevant to $B$.

$$p(B \mid A) = p(B)$$

But Stalnaker's Thesis was supposed to hold for many pairs of propositions $A, B$.

So, if the derivation is sound, then Stalnaker's Thesis can hold generally only in trivial probability spaces.

This is why Lewis's result is not just a problem about one conditional. It is a problem about the whole project of identifying:

probability of a conditional with conditional probability.

# The connection

# DAB as a conditional thesis

Those remarks give us a hint about how we can make the connection between the DAB argument, and the triviality result for Stalnaker's Thesis even closer.

Those remarks give us a hint about how we can make the connection between the DAB argument, and the triviality result for Stalnaker's Thesis even closer.

Let $G$ be the proposition that things are good. Then DAB can be written:

**Conditional DAB**

$$V(A) = p(A \to G).$$

Those remarks give us a hint about how we can make the connection between the DAB argument, and the triviality result for Stalnaker's Thesis even closer.

Let $G$ be the proposition that things are good. Then DAB can be written:

**Conditional DAB**

$$V(A) = p(A \to G).$$

In words:

Your desire for $A$ should match your credence in
*If $A$, then things are/would be good.*

# Jeffrey plus DAB yields Stalnaker

Recall the binary-goodness version of Jeffrey's theory:

$$V(A) = p(G \mid A).$$

# Jeffrey plus DAB yields Stalnaker

Recall the binary-goodness version of Jeffrey's theory:

$$V(A) = p(G \mid A).$$

And conditional DAB says:

$$V(A) = p(A \to G).$$

## Jeffrey plus DAB yields Stalnaker

Recall the binary-goodness version of Jeffrey's theory:

$$V(A) = p(G \mid A).$$

And conditional DAB says:

$$V(A) = p(A \to G).$$

Putting them together:

**The bridge**

$$p(A \to G) = p(G \mid A).$$

# Jeffrey plus DAB yields Stalnaker

Recall the binary-goodness version of Jeffrey's theory:

$$V(A) = p(G \mid A).$$

And conditional DAB says:

$$V(A) = p(A \to G).$$

Putting them together:

**The bridge**

$$p(A \to G) = p(G \mid A).$$

But this is just Stalnaker's Thesis, with $G$ as consequent.

# The shared structure

| | DAB | Stalnaker |
|---|---|---|
| Thesis | $V(A) = p(A^\circ)$ | $p(A \rightarrow B) = p(B \mid A)$ |
| Problematic move | $A^\circ$ is fixed across learning | $A \rightarrow B$ is fixed across learning |
| Result | $p(A^\circ) = p(A^\circ \mid A)$ | $p(B) = p(B \mid A)$ |

# Conclusions and Options

At the outset, I said that DAB is very popular in ethics.

## Conclusions and Options

At the outset, I said that DAB is very popular in ethics.

It would be very surprising, then, if a certain decision-theoretic view was inconsistent with it.

At the outset, I said that DAB is very popular in ethics.

It would be very surprising, then, if a certain decision-theoretic view was inconsistent with it.

Resistance strategies:

# Conclusions and Options

At the outset, I said that DAB is very popular in ethics.

It would be very surprising, then, if a certain decision-theoretic view was inconsistent with it.

Resistance strategies:

- Context dependence?

# Conclusions and Options

At the outset, I said that DAB is very popular in ethics.

It would be very surprising, then, if a certain decision-theoretic view was inconsistent with it.

Resistance strategies:

- Context dependence?
- Giving up CRT.

# Conclusions and Options

At the outset, I said that DAB is very popular in ethics.

It would be very surprising, then, if a certain decision-theoretic view was inconsistent with it.

Resistance strategies:

- Context dependence?
- Giving up CRT.
- Giving up Closure.

# Conclusions and Options

At the outset, I said that DAB is very popular in ethics.

It would be very surprising, then, if a certain decision-theoretic view was inconsistent with it.

Resistance strategies:

- Context dependence?
- Giving up CRT.
- Giving up Closure.
- Giving up the Principal Principle (and switching to a similar principle, i.e. the new Principle).

# Conclusions and Options

At the outset, I said that DAB is very popular in ethics.

It would be very surprising, then, if a certain decision-theoretic view was inconsistent with it.

Resistance strategies:

- Context dependence?
- Giving up CRT.
- Giving up Closure.
- Giving up the Principal Principle (and switching to a similar principle, i.e. the new Principle).

# References


James M. Joyce. *The Foundations of Causal Decision Theory*. Cambridge University Press, 1999.

David Lewis. Probabilities of conditionals and conditional probabilities. *The Philosophical Review*, 85(3):297, Jul 1976. ISSN 0031-8108. doi: 10.2307/2184045. URL `http://dx.doi.org/10.2307/2184045`.

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`.

David Lewis. Desire as belief. *Mind*, XCVII(387):323–332, 1988. ISSN 1460-2113. doi: 10.1093/mind/xcvii.387.323. URL `http://dx.doi.org/10.1093/mind/XCVII.387.323`.

David Lewis. Desire as belief II. *Mind*, 105(418):303–313, Apr 1996. ISSN 1460-2113. doi: 10.1093/mind/105.418.303. URL `http://dx.doi.org/10.1093/mind/105.418.303`.

Derek Parfit. *Reasons and Persons*. Oxford University Press, 1984.

Robert Stalnaker. Probability and conditionals. *Philosophy of Science*, 37(1):64–80, Mar 1970. ISSN 1539-767X. doi: 10.1086/288280. URL `http://dx.doi.org/10.1086/288280`.