probability and counterfactuals

fabrizio cariani and paolo santorio

northwestern + maryland

august 8, 2019

Lecture 4

recap of L3: indicative triviality

- the thesis seems plausible, but...
- ... together with extra assumptions it entails "triviality"
- the probability function is trivial because it is forced to have $p(\mathrm{B} \mid \mathrm{A}) = p(\mathrm{B})$ for almost any A and B; that's bad :(
- maybe the thesis is false or maybe it's the extra assumptions that are the problem

recap of L3: setup for counterfactual triviality

- we set up Skyrms' thesis and CPP by analogy with Stalnaker's thesis
- are they trivial too?
- *good news:* the exact same argument doesn't work
- *bad news:* a similar argument relying on different extra assumptions does work

today's plan

1. one more spin on the indicative triviality merry-go-round
2. the counterfactual triviality result
3. a response to the counterfactual triviality result
4. a skeptical take on probabilities of conditional from linguistics

an example

The Purple Party and the Yellow Party are competing for the election. If elected, the Purple Party will raise wages and double healthcare funds. If elected, the Yellow Party will cut taxes and restrict immigration. The Purple Party is far ahead in the polls, but it's not quite certain that they will win.

consider the following two claims:

(1) wages will be raised. **raise**

(2) if Yellow wins, wages will be raised. **yellow→raise**

some obvious intuitions:

$$c(\textbf{raise}) = \text{high} \qquad c(\textbf{yellow}{\rightarrow}\textbf{raise}) = \text{low}$$

an example

in the background:

- standard probability theory
- conditionalization is the rational update rule

$$P_E(A) = P(A \mid E)$$

- **closure**: if you are rational and you conditionalize on any claim, you stay rational

an example

Bradley-style premises for this case:

(a) If $c(\textbf{raise}) = 1$, then $c(\textbf{yellow} \rightarrow \textbf{raise}) = 1$

(b) If $c(\textbf{raise}) = 0$, then $c(\textbf{yellow} \rightarrow \textbf{raise}) = 0$

and relatedly, via **closure**:

(a′) $c_{\textbf{raise}}(\textbf{yellow} \rightarrow \textbf{raise}) = 1$
(if you learn **raise**, you should be certain of **yellow**→**raise**)

(b′) $c_{\neg\textbf{raise}}(\textbf{yellow} \rightarrow \textbf{raise}) = 0$
(if you learn ¬**raise**, you should be certain that **yellow**→**raise** is false)

an example

the triviality argument for this case

1. $c(\textbf{yellow} \to \textbf{raise}) =$
2. $c((\textbf{yellow} \to \textbf{raise}) \wedge \textbf{raise}) + c((\textbf{yellow} \to \textbf{raise}) \wedge \neg\textbf{raise}) =$
3. $c((\textbf{yellow} \to \textbf{raise}) \mid \textbf{raise}) \times c(\textbf{raise}) +$
   $c((\textbf{yellow} \to \textbf{raise}) \mid \neg\textbf{raise}) \times c(\neg\textbf{raise}) =$
4. $1 \times c(\textbf{raise}) + 0 \times c(\neg\textbf{raise}) =$
5. $Pr(\textbf{raise})$

resisting triviality

triviality argument tells us that one of the following needs to go:

1. standard probability theory
2. conditionalization is the rational rule for updating credences
3. **closure** (you can conditionalize on any proposition)
4. Bradley-style principles:
    (a) If $c(\textbf{raise}) = 1$, then $c(\textbf{yellow} \rightarrow \textbf{raise}) = 1$
    (b) If $c(\textbf{raise}) = 0$, then $c(\textbf{yellow} \rightarrow \textbf{raise}) = 0$
5. $c(\textbf{raise}) > c(\textbf{yellow} \rightarrow \textbf{raise})$

which one? fill in the survey to tell us!

resisting triviality

triviality argument tells us that one of the following needs to go:

1. standard probability theory
2. conditionalization is the rational rule for updating credences
3. **closure** (you can conditionalize on any proposition)
4. Bradley-style principles:
   (a) If $c(\textbf{raise}) = 1$, then $c(\textbf{yellow} \rightarrow \textbf{raise}) = 1$
   (b) If $c(\textbf{raise}) = 0$, then $c(\textbf{yellow} \rightarrow \textbf{raise}) = 0$
5. $c(\textbf{raise}) > c(\textbf{yellow}{\rightarrow}\textbf{raise})$

which one? fill in the survey to tell us!

resisting triviality

triviality argument tells us that one of the following needs to go:

1. standard probability theory
2. conditionalization is the rational rule for updating credences
3. **closure** (you can conditionalize on any proposition)
4. Bradley-style principles:
   (a) If $c(\textbf{raise}) = 1$, then $c(\textbf{yellow} \rightarrow \textbf{raise}) = 1$
   (b) If $c(\textbf{raise}) = 0$, then $c(\textbf{yellow} \rightarrow \textbf{raise}) = 0$
5. $c(\textbf{raise}) > c(\textbf{yellow} {\rightarrow} \textbf{raise})$

which one? fill in the survey to tell us!

resisting triviality

triviality argument tells us that one of the following needs to go:

1. standard probability theory
2. conditionalization is the rational rule for updating credences
3. **closure** (you can conditionalize on any proposition)
4. Bradley-style principles:
   (a) If $c(\textbf{raise}) = 1$, then $c(\textbf{yellow} \rightarrow \textbf{raise}) = 1$
   (b) If $c(\textbf{raise}) = 0$, then $c(\textbf{yellow} \rightarrow \textbf{raise}) = 0$
5. $c(\textbf{raise}) > c(\textbf{yellow} \rightarrow \textbf{raise})$

which one? fill in the survey to tell us!

resisting triviality

triviality argument tells us that one of the following needs to go:

1. standard probability theory
2. conditionalization is the rational rule for updating credences
3. **closure** (you can conditionalize on any proposition)
4. Bradley-style principles:
   (a) If $c(\textbf{raise}) = 1$, then $c(\textbf{yellow} \rightarrow \textbf{raise}) = 1$
   (b) If $c(\textbf{raise}) = 0$, then $c(\textbf{yellow} \rightarrow \textbf{raise}) = 0$
5. $c(\textbf{raise}) > c(\textbf{yellow} \rightarrow \textbf{raise})$

which one? fill in the survey to tell us!

resisting triviality

triviality argument tells us that one of the following needs to go:

1. standard probability theory
2. conditionalization is the rational rule for updating credences
3. **closure** (you can conditionalize on any proposition)
4. Bradley-style principles:
   (a) If $c(\textbf{raise}) = 1$, then $c(\textbf{yellow} \rightarrow \textbf{raise}) = 1$
   (b) If $c(\textbf{raise}) = 0$, then $c(\textbf{yellow} \rightarrow \textbf{raise}) = 0$
5. $c(\textbf{raise}) > c(\textbf{yellow}{\rightarrow}\textbf{raise})$

which one? fill in the survey to tell us!

resisting triviality

triviality argument tells us that one of the following needs to go:

1. standard probability theory
2. conditionalization is the rational rule for updating credences
3. **closure** (you can conditionalize on any proposition)
4. Bradley-style principles:
    (a) If $c(\textbf{raise}) = 1$, then $c(\textbf{yellow} \rightarrow \textbf{raise}) = 1$
    (b) If $c(\textbf{raise}) = 0$, then $c(\textbf{yellow} \rightarrow \textbf{raise}) = 0$
5. $c(\textbf{raise}) > c(\textbf{yellow}{\rightarrow}\textbf{raise})$

which one? fill in the survey to tell us!

back to counterfactual triviality: Williams′ proof

the strategy for counterfactual triviality

Williams′ idea: nothing in triviality proofs relies on the semantics for indicatives; so we can reproduce them for counterfactuals!

The overall strategy:

i. start from a version of Skyrms′ Thesis

ii. derive a Thesis-like principle about chance

iii. motivate a closure assumption about chance

iv. run a Lewis-like triviality result using the above

norms for counterfactual supposition

starting point: an attitude of counterfactual supposition.

> *supposition is a kind of mental activity familiar to all of us. We might suppose that our train won't arrive in order to form a contingency plan for that eventuality, and believe-under-that-supposition that the best thing to do is to take a cab.*

norms for counterfactual supposition

an assumption: there is a normative connection between this attitude and beliefs in conditionals:

counterfactual Ramsey test (CRT). $c_{w,t}(A \mathrel{\Box\!\!\rightarrow} B) = c^{A}_{w,t}(B)$

> *what a Ramsey identity asserts is a normative connection between two distinct mental states: for fully rational agents, the degree of suppositional belief in B on A and the corresponding categorical credence in 'if A then B' (/'if were A then B') should coincide. It is perfectly possible for agents to have degrees of belief in conditionals that diverge from the corresponding suppositional credences—but if the Ramsey Identities are correct, this is a form of irrationality.*

norms for counterfactual supposition

incidentally: weaker versions of CRT will also create trouble:

CRB. $c(A \boxright B) \leq c^A(B)$
CRZ. If $c^A(B) = 0$, then $c(A \boxright B) = 0$

we focus on the triviality argument we gets from CRT, but we could run analogous proofs with CRB and CRZ

skyrms′ thesis, again

a further assumption: a version of Skyrms′ Thesis

$$c^{\mathsf{A}}(\mathsf{B}) = \sum_{ch \in \mathbf{CH}} c[\chi(ch)] \cdot ch(\mathsf{B} \mid \mathsf{A})$$

- for the right-hand side, we are using our version of Skyrms′ Thesis from L2
- for the left-hand side, we are using Williams′ ‘suppositional probabilities′

informed skyrms' thesis

- assume: we have an agent who is ideally informed about the chance function
- so, for this agent, Skyrms' Thesis reduces to the simple equation:

$$\textbf{Informed Skyrms' Thesis.} \quad c^{A}_{w,t}(B) = ch_{w,t}(B \mid A)$$

- we're going to use this as a premise in the triviality proof

principal principle for informed agents

recall the Principal Principle:

$$\textbf{Principal Principle.} \quad ic(\mathsf{A} \mid H_t^w \;\&\; T^w) = ch_{w,t}(\mathsf{A})$$

- suppose that, in addition to being ideally informed about chances, at t our agent has learned with certainty all history up to t (i.e. she has learned $H_t^w \;\&\; T^w$)
- then via the Principal Principle, her credences at t are constrained as follows:

$$c_{w,t}(\mathsf{A}) = ch_{w,t}(\mathsf{A})$$

- in particular, we get the instance:

$$c_{w,t}(\mathsf{A} \mathrel{\Box\!\!\rightarrow} \mathsf{B}) = ch_{w,t}(\mathsf{A} \mathrel{\Box\!\!\rightarrow} \mathsf{B})$$

the chancy equation

now, putting together the three equations:

$$c_{w,t}(A \,\Box\!\!\rightarrow B) = c^{A}_{w,t}(B)$$
$$c^{A}_{w,t}(B) = ch_{w,t}(B \mid A)$$
$$c_{w,t}(A \,\Box\!\!\rightarrow B) = ch_{w,t}(A \,\Box\!\!\rightarrow B)$$

we get the following:

Chancy Equation. $ch_{w,t}(B \mid A) = ch_{w,t}(A \,\Box\!\!\rightarrow B)$

what does this remind you of?

a chance equivalent of the Thesis!

(from now on, we'll drop the subscripts)

the chancy equation

now, putting together the three equations:

$$c_{w,t}(A \mathrel{\Box\!\!\rightarrow} B) = c^{A}_{w,t}(B)$$
$$c^{A}_{w,t}(B) = ch_{w,t}(B \mid A)$$
$$c_{w,t}(A \mathrel{\Box\!\!\rightarrow} B) = ch_{w,t}(A \mathrel{\Box\!\!\rightarrow} B)$$

we get the following:

Chancy Equation. $\quad ch_{w,t}(B \mid A) = ch_{w,t}(A \mathrel{\Box\!\!\rightarrow} B)$

what does this remind you of?

a chance equivalent of the Thesis!

(from now on, we'll drop the subscripts)

closure for chances

we need one more assumption: a **closure** principle for chances

closure for chances (CfC)
given that *ch* is a chance distribution at t, and given an arbitrary sentence X, there are worlds w', w'' and times t' and t'' such that $ch_X = ch(\bullet \mid X)$ is the objective chance distribution at w', t' and $ch_{\bar{X}} = ch(\bullet \mid \bar{X})$ likewise at w'', t''

- notice: CfC only says that, given that there *Ch* is the chance function of w at t, there is some world or other such that Ch_X is the chance function (not that this second world has to be close/similar to w in any way)

closure for chances

why believe CfC?

- a standard assumption about chances (defended by Lewis himself): chances evolve by conditionalization
- i.e.: chances at a given time arise from earlier chance distributions by conditionalizing on what has in fact occurred in the intervening period

maybe this won't give us a full closure principle. but, as long as closure holds in some cases, we will be able to run some instances of the triviality argument!

re-running the lewis proof

now we run an argument analogous to lewis's:

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

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

iii. $ch_C(A \mathrel{\Box\!\!\rightarrow} C) \times ch(C) + ch_{\bar{C}}(A \mathrel{\Box\!\!\rightarrow} C) \times ch(\neg C) =$

iv. $ch_C(C \mid A) \times ch(C) + ch_{\bar{C}}(C \mid A) \times ch(\neg C) =$

v. $1 \times ch(C) + 0 \times ch(\neg C) =$

vi. $ch(C)$

This result is as bad as the previous one!

resistance strategies

- challenging **closure** (Williams′ preferred way out)
- challenging some of the assumptions, including CRT and PP
- deny that counterfactuals have chances
- (fancy) argue that the PP needs to be restricted

Schwarz's (2016) reply

there is a problem with Williams's argument that has nothing to do with the probabilities of counterfactuals.

Schwarz's (2016) reply: preliminaries

observation one: because chances are objective features of the world, chance claims themselves have chances

Schwartz's (2016) reply: preliminaries

observation two: according to PP, every chance function must think of itself that it has chance 1 ("is self-aware").

- $\chi(ch) =$ the proposition that is true in world w iff ch is the true chance function in w (suppose that ch governs all times in w)
- now skipping some steps, PP will entail this:

$$c(\chi(ch) \mid \chi(ch)) = ch_w(\chi(ch))$$

- but $c(\chi(ch) \mid \chi(ch)) = 1$
- so, ch_w must think of itself that it has chance 1

Schwarz's (2016) reply: core point

observation three: the idea that chance functions must be self-aware is consistent on its own; but it's inconsistent with closure for chance functions.

- let A be any proposition that has intermediate chance according to ch (i.e. $0 < ch(\mathsf{A}) < 1$)
- consider the conditionalized ch_{A} applied to $\chi(ch)$

$$ch_{\mathsf{A}}(\chi(ch)) = ch(\chi(ch) \mid \mathsf{A})$$

- but because $ch(\chi(ch)) = 1$, $ch(\chi(ch) \mid \mathsf{A}) = 1$, so:

$$1 = ch_{\mathsf{A}}(\chi(ch)) \neq ch_{\mathsf{A}}(\chi(ch_{\mathsf{A}})) = 1$$

Schwarz's (2016) reply: proposed fix

- PP is too strong;
- so, fall back on a principle without self-awareness
- this exists in the literature: The New Principle
- see Hall (1994); Lewis (1994)

domain restriction

rethinking the dialectic

(I) intuitions about probabilities of conditionals

(T) Stalnaker's thesis

(E) extra Stuff

our reasoning:

- (I) supports (T)
- however, (T)+(E) yields triviality :(
- we have to backtrack

identifying a strategy

(I) intuitions about probabilities of conditionals

(T) Stalnaker's thesis

(E) extra Stuff

backtracking options:

- deny (I) but *how*?
- deny (E)
- deny that (I) supports (T) by providing an account of (I) that supports (I) but does not support (T)

domain restriction: Lewisian prehistory

(3) Always/Usually/Sometimes, if Vera drinks milk, she spills some of it

- in (3), "if" is not really a conditional...
- ... its job is to restrict the adverbs of quantification

domain restriction: Kratzer

(4) if she is dancing, she must be happy

- *if* restricts the modal *must* ...
- ... just like Lewis thought in the case of AOQ sentences
- the logical form of (4) is as in (5)

(5) (*must*: she is happy) (she is dancing)

restriction on probability operators

(6) if she is dancing, she is probably happy

(7) probably, if she is dancing, she is happy

(8) (probably: she is dancing) (she is happy)

restriction on probability operators

(9) if she is dancing, she is happy with probability 0.7

(10) probably, if she is dancing, she is happy with probability 0.7

(11) (with-probability-0.7: she is dancing) (she is happy)

an extra look at the argument structure

The flat-footed analysis:

- with-probability-0.7 [*if* she is dancing, she is happy]

The domain restriction analysis:

- (with-probability-0.7 (*if* she is dancing)) (she is happy)

the basic point

we sometimes say that the probability of (12) is 0.7

(12) if she is dancing, she is happy

but that's not what we mean

Instead, what we mean is that the sentence (13) is true

(13) if she is dancing, with probability 0.7 she is happy

(14) (with-probability-0.7 (*if* she is dancing)) (she is happy)

the effect on the dialectic

the intuition we used to justify the Thesis was that it predicts our judgments about the probabilities of conditionals

the restriction theorist says:

> *those motivating judgments were not to the effect that the conditional has probability 0.7;*
> *instead they are to the effect that the restricted sentence is true*

three arguments against domain restriction

general arguments against restriction 1

probabilities can be ascribed to conditionals *via* propositional anaphora, as in:

> *Elsa: if the cats fell in the pool, they didn't like it*
> *Anna: That's (60 %) likely*

Elsa: if the cats fell in the pool, they didn't like it
Anna: That's (60 %) likely

Argument:

P1. Anna's "That" refers to the proposition expressed by Elsa's conditional utterance

P2. Anna ascribes probability to whatever proposition Elsa expresses

C. Anna ascribes probability to the proposition expressed by Elsa

general arguments against restriction 2

> Coins. *Martina is considering tossing two fair coins, A and B, in two independent tosses. You leave the room before you discover whether she tosses them or not.*

now, assess the probability of:

(15) coin A landed heads, if it was tossed, and coin B landed tails, if it was tossed.

(16) each of coin A and coin B landed heads, if it was tossed.

(17) with probability 1/4, coin A landed heads, if it was tossed, and coin B landed tails, if it was tossed.

(18) with probability 1/4, each of coin A and coin B landed heads, if it was tossed.

specifically counterfactual arguments

some objection attack the domain restriction strategy in the counterfactual case specifically

in *Counterfactuals and Probability*, Schulz objects that the domain restriction approach does not do well with:

(19) It is 90% likely that no one else would have killed Kennedy if Oswald had not done it

on the domain restriction approach, this becomes:

(20) (with-probability-0.9 (if Oswald had not done it)) (no one else would have killed Kennedy)

but with-probability-0.9 is an epistemic probability operator, expressing something like indicative supposition

words of wisdom (from Schulz)

indicative and counterfactual 'if'-clauses seem to invite us to do different things. An indicative conditional is evaluated by temporarily adding the ante- cedent to our stock of beliefs and then asking whether the consequent is true in this more informed belief state. There is a natural connection with the restrictor view here. [...] Counterfactual 'if '-clauses also invite us to make a supposition, but a counterfactual supposition is different from an indicative one. Making a counterfactual supposition often requires us to temporarily suspend certain beliefs while holding others constant.