Triviality and Tenability

Day 5 // ESSLLI, Conditionals and Information-sensitivity // David Boylan and Matt Mandelkern

1 The Thesis

Generalisation: seems like your probability in a conditional should just be the corresponding conditional probability.

- *The Thesis*: For all a, c, Pr : $Pr(a > c) = Pr(c|a)$ if $Pr(a) > 0$
  - Say that *Pr yields the Thesis* for a conditional operator $>$ iff Pr and $>$ satisfy the above equation.

Due to Adams 1975; Stalnaker 1970. Sometimes called 'Adams' Thesis', 'Stalnaker's Thesis', or 'The Equation'.

- Could be thought of as an extra rational constraint analogous to Normality and Additivity; though some, particularly "non-factualists", think it is a constitutive fact, not a normative ideal.

A natural thought is that the Thesis doesn't just hold as a one-off; it continues to hold as you learn more information.

- Say that the Thesis holds across conditionalisation (wrt Pr and $>$) iff $Pr(\cdot|E)$ yields the Thesis wrt $>$, whenever Pr does.

Assuming $Pr(\cdot|E)$ is defined.

- This would predict a probability version of Import-Export:
  - $Pr(A > (B > C)) = Pr((A \wedge B) > C)$
- It also predicts the failures of MP we saw:
  - Simply let $Pr(Reagan) = 0.7$, $Pr(Dem) = 0.25$ and $Pr(Anderson) = 0.05$

Dynamic triviality arguments show this natural thought leads to disaster.

Another class of triviality results show that the Thesis has to be restricted, even when considering a single probability function; it must be restricted to antecedents that do not contain conditionals.
(Lewis, 1976)

2 Triviality 1

Assumptions:

1. All rational probability functions yield the Thesis.
2. The class of rational probability functions is closed under conditionalisation: if Pr is rational, then so too is Pr_E.
   - i.e. conditionalisation is the right rule for updating our probabilities, when we get new information.
3. $0 < Pr(c) < 1$ and $Pr(ac) > 0$ and $Pr(a\neg c) > 0$,

Now we have:

$$Pr(a > c) = Pr(a > c|c)Pr(c) + Pr(a > c|\neg c)Pr(\neg c)$$

by the law of total probability. $Pr(b|a) = \frac{Pr(ab)}{Pr(a)}$

$$Pr(a > c|c) = Pr_c(a > c) = Pr_c(c|a) = 1$$

by *The Thesis*, where Pr_c is obtained from Pr by conditioning on c

$$Pr(a > c|\neg c) = Pr_{\neg c}(a > c) = Pr_{\neg c}(c|a) = 0$$

by *The Thesis*, where $Pr_{\neg c}$ is obtained from Pr by conditioning on $\neg c$

$$Pr(a > c) = \underbrace{Pr(a > c|c)}_{1} Pr(c) + \underbrace{Pr(a > c|\neg c)}_{0} Pr(\neg c) = Pr(c)$$

But it's implausible that the probability of a conditional equals the probability of its consequent:

(1) If the car crashes, the airbag will go off.

3 Triviality 2

Assumptions:

- All rational probability functions yield the Thesis.
- For some rational Pr, $0 < Pr(c) < 1$ and $Pr(ac) > 0$ and $Pr(a\neg c) > 0$,
- For that same Pr, $Pr_{a\vee(\neg a\wedge a>c)}$ is rational.

The argument informally:

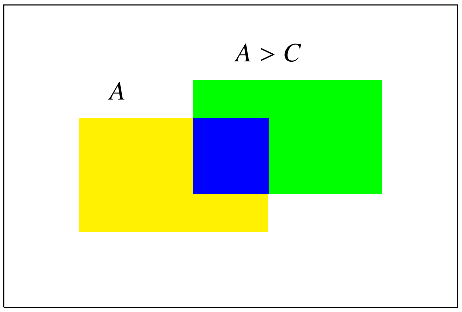
- Since Pr is rational, $Pr(a > c) = Pr(c|a)$
- Now consider$Pr_{a\vee(\neg a\wedge a>c)}$.
  - $Pr_{a\vee(\neg a\wedge a>c)}(c|a)$ cannot change: it must be $Pr(c|a)$.
  - $Pr_{a\vee(\neg a\wedge a>c)}(a > c)$ *must* change: since we discard worlds where $a > c$ is false, but not worlds where it's true.
- So the Thesis does not in fact hold on $Pr_{a\vee(\neg a\wedge a>c)}$. Contradiction.

The argument formally:

1. $Pr_{a\vee(\neg a\wedge a>c)}(a > c) = Pr(a > c|a)Pr(a|a \vee (\neg a \wedge a > c)) +$
$$Pr(a > c|\neg a \wedge a > c)Pr(\neg a \wedge a > c|a \vee (\neg a \wedge a > c))$$
2. $Pr(a > c|a) = \frac{Pr(a\wedge a>c)}{Pr(a)} = \frac{Pr(ac)}{Pr(a)} = Pr(c|a)$ by strong centering

$$Pr(a > c|\neg a \wedge a > c) = 1, \text{ so}$$

Logical Space



$Pr_{a \vee (\neg a \wedge a > c)}(a > c) = Pr(c|a)Pr(a|a \vee (\neg a \wedge a > c)) + Pr(\neg a \wedge a > c|a \vee (\neg a \wedge a > c)) \geq$

$Pr(c|a)Pr(a|a \vee (\neg a \wedge a > c)) + Pr(c|a)Pr(\neg a \wedge a > c|a \vee (\neg a \wedge a > c)) = Pr(c|a)$.

But $Pr_{a \vee (\neg a \wedge a > c)}(c|a) = Pr(c|a)$.

So if we start with a Thesis-friendly probability measure, we can easily get to a Thesis-unfriendly one, just by conditionalizing on $a \vee (\neg a \wedge a > c)$.

4 Triviality 3

Bradley notes that The Thesis entails

Preservation: $Pr(c) = 0 \wedge Pr(a) > 0 \rightarrow Pr(a > c) = 0$.

Assumptions:

1. Same as before.
2. If A is consistent, then there is some reasonable probability function such that $Pr(A) = 1$

Argument:

- From 1 we get that *Preservation* holds for all probability measures.
- Then either a or $a > c$ must entail c.
  - Otherwise, there will be parts of the state-space where $(a > c) \wedge \bar{c}$ holds and likewise parts where $a\bar{c}$ holds.
  - Assign the disjunction of these states probability 1 and we'll have a counterexample to *Preservation*.

5 What to say?

Two options:

1. Reject the claim that rational probability functions are closed under conditionalisation.
   - Radical — this is a core part of orthodox Bayesianism.
   - Notice as well we must deny it in some very simple cases: there will be cases where Pr is rational, but $Pr(\cdot|A)$ is not!
   - Not immediately obvious how it helps with Preservation argument.
2. Restrict the range of measures which yield the Thesis somehow.

First reason for option 2: Thesis imposes a strong constraint.

(Ellis, 1978)

- Note that The Thesis holding is equivalent to a conditional being probabilistically independent of its antecedent, given Strong Centering:

$$Pr(a > c) = Pr(c|a) \equiv$$
$$Pr(a > c) = \frac{Pr(ac)}{Pr(a)} \equiv$$
$$Pr(a > c) = \frac{Pr(a \wedge (a > c))}{Pr(a)} \equiv$$
$$Pr(a > c) = Pr(a > c|a)$$

- As Stalnaker (1974) pointed out, once we see that, it's not really clear why we would expect The Thesis to hold in general, or why we would want it to.
- Not clear that this restricts the Thesis enough to avoid the triviality arguments.

That raises the question: can we construct cases where the conditional is intuitively not probabilistically dependent of its antecedent? Something we'll return to.

A different reason to restrict: *context-sensitivity*.

- We've seen abundant reason to think the sentence ⌜a > c ⌝ is *context sensitive*.
  - In a context where the relevant information is i it might express the proposition $A >_i C$; if the relevant information is some other set of worlds i' it will express some potentially distinct proposition $A >_{i'} C$
- Once we make this distinction, we are forced to make some choices.
  - Given a particular conditional $>_i$, which rational probability measures yield the Thesis for $>_i$?

  - Likewise, given a particular probability measure Pr, which conditionals $>_i$ yield the Thesis for Pr?

A natural idea: the conditional should be coordinated with your evidence:

- Let E_{Pr} be the strongest proposition such that $Pr(E) = 1$.
- Assume that for each rational probability function, $>_{E_{Pr}}$ yields Stalnaker's Thesis wrt to Pr. Call this assumption *Locality*.
  - Roughly, coordinate your conditional probabilities with *your* conditional.
- Seems pretty intuitive:
  - Suppose E_A is Alice's total evidence and E_B is Billy's.
  - Why should Alice have to coordinate her probabilities in $A >_{E_B} C$ — Billy's conditional — with her conditional probabilities?

6 Back to the Triviality Arguments

Response to Triviality 1:

- Assumption 1 fails: there's no particular conditional which yields Stalnaker's Thesis on all reasonable probability functions.
- Particular step where the argument fails:
  - Suppose Pr satisfies the other assumptions.
    - Again, say that E_{Pr} is the strongest proposition such that $Pr(E) = 1$.
  - Locality says that $Pr(A >_E C) = Pr(C|A)$.
  - It does not follow from Locality that $Pr_C(A >_E C) = Pr(C|A)$
    - Instead Pr_C yields the Thesis for a *different* conditional, namely $A >_{E \cap C} C$.

Triviality 1 clearly fails then. The same kind of argument applies to Triviality 4 and arguably to Triviality 3.

7 Tenability in finite cases

Basic idea for the construction:

Presentation here is heavily indebted to Justin Khoo, Paolo Santorio and Simon Goldstein.

- We start with a probability function over a state space over worlds.
- We extend that probability function to one over a state space of world-selection function pairs.
- Conditional propositions are constructed using the selection functions; intuitively they rule out certain selection functions.

- Probabilities of non-conditional propositions remain the same in the new state space.

Selection functions and sequences:

- Stalnakerian selection functions determine a total order over worlds:
  - given a world w for any two distinct worlds, one must be strictly closer to the w.
- This order can be thought of as a sequence $\langle w_1, w_2, ... \rangle$, where w_1 is the closest world to w_1, i.e. w_1 itself, w_2 is the next closest world, w_3 is the next closest world, and so on.
- So a pair $\langle w, f \rangle$ can instead be thought of as a sequence beginning with w, where w_i precedes w_j in the sequence iff $w_i <_w^f w_j$.

Well, this is only really right when the selection function has special properties. Luckily, those properties are plausible: it is the same conditional that we need to validate Flattening, which we saw on Thursday.

Definition of a new state space, given an old *finite* state space W:

- W^*, the set of "worlds", is the set of sequences of worlds in W.
- Set of propositions is just the powerset of W^*.
- Any old proposition A in the old algebra has a representative in the new one: the set of sequences starting with a world in A.
  - We'll write this as A^*
- The conditional $A > C$ is true at a sequence s iff the first A-world in s is a C-world.

Note this definition does not allow left or right-nested conditionals. To generalise we need a more complicated definition. Where π is a sequence, say that $\pi[n]$ is the sequence that begins with π's n member

New probability function Pr^* over W^*:

- Probability of a sequence is the probability of "drawing" a sequence of worlds *without replacement*.
  - E.g. $Pr(\langle w_1, w_2, w_3, w_4 \rangle) = Pr(w_1) \times Pr(w_2 | \neg w_1) \times Pr(w_3 | \neg(w_1 \vee w_2)) \times Pr(w_4 | \neg(w_1 \vee w_2 \vee w_3))$
- Probability of a set of sequences obtained by adding probabilities of the individual sequences.

More precisely:

- Say that $[w_1, ..., w_n]$ is the set of sequences starting with w_1, ..., w_n
- Recursive definition of Pr^*
  - $Pr^*([w_1]) = Pr(w_1)$
  - $Pr^*([w_1, ..., w_{n-1}, w_n]) = Pr^*([w_1, ..., w_{n-1}]) \times \frac{Pr(w_n)}{Pr(W - w_1, ..., w_{n-1})}$

Thesis will hold in such models for non-conditional antecedents and consequents. Simple case:

- Three worlds: coin *N*ever tossed, coin lands *H*eads, coin lands *T*ails.
- $Pr(N) = 1/2$; $Pr(H) = 1/4$; $Pr(T) = 1/4$

8 Tenability with Information-Sensitivity

Two related issues with this simple construction. Issue 1: where's the information-sensitivity?

- Information-sensitivity was supposed to block triviality.

  But no apparent info sensitivity in these models.

Issue 2: updating.

- Suppose we conditionalise on $H \vee T$. Resulting model no longer validates the Thesis.
  - Intuitively the following claim should get 0:

    (2) If the coin didn't land heads, it wasn't flipped.

    But in fact it gets prob 1/4!

To add information-sensitivity, suppose we have a set of conditional operators, $>_A$, $>_B$, ... , each indexed to a proposition.

- $A >_E C$ is true at a sequence s iff the first $A \cap E$-world in the sequence is a C-world.

Now each *Pr* has a conditional which yields the thesis for some operator:

- For non-conditional A and C, $Pr^*(A >_{E_{Pr}} C) = Pr^*(C|A)$, when $Pr(A) \neq 0$.

Recall that E_{Pr} is the strongest proposition such that $Pr(E) = 1$.

9 Generalising Tenability

We've ended up, for independent reasons, with a model much like the one that van Fraassen 1976 developed to give a tenability result for the intra-contextual version of The Thesis for conditionals with Boolean antecedents. VF's approach, roughly was to extend a pmf p from worlds to vectors, roughly as follows:

working with a discrete model for simplicity; vF's is continuous.

- start with an ur-pmf p; we get the probability associated with a context c of φ by conditioning p on the context set κ_c.

- $[\langle w_1, \ldots w_n \rangle]$ is the set of c-admissible vectors which start with the subsequence $\langle w_1, \ldots w_n \rangle$.
- $\langle w_1, \ldots w_n, \_ \rangle$ is the set of worlds that follow the sequence $\langle w_1, \ldots w_n \rangle$ in some contextually admissible vector.
- $p([\langle w, w_1, w_2, \ldots w_n \rangle]) = 0$ if $[\langle w, w_1, w_2, \ldots w_n \rangle] = \varnothing$, otherwise:

$$p([w]) = p(w)$$
$$p([w_1, \ldots w_n, w_{n+1}]) = p([w_1 \ldots w_n]) \cdot \frac{p(w_{n+1})}{p(\langle w_1, \ldots w_n, \_ \rangle)}$$

vF calls these 'Stalnaker Bernoulli models'. The intuition is that the probability that the n^{th} world is w is always $\frac{p(w)}{p(\mathbf{n})}$, where $\mathbf{n}$ is the set of worlds in any n^{th} position.

The logical motivations for models like this from logic provide motivation to take them seriously, not just as tenability models.

And the perspective that bounds help us zero in on a model help, in turn, make sense of some striking failures of The Thesis.

Again, we assume that we default to maximal vector models, where a vector model for p in context c is maximal just in case, if we add any more vectors to the model, p will not be satt in c.

10 Failures of The Vanilla Thesis

If we use the vF construction without bounds (or restricted accessibility), we get the intracontextual, vanilla-antecedent thesis. But there are some striking generalizations that that misses, which our bound-guided model captures. hence, TVT

10.1 Complex conditionals

The anti-MP, pro-IE intuitions we have looked at also show up in probability judgments. Recall:

(3) If a Republican wins, then if Reagan doesn't, Anderson will.

Intuitively, your credence in (3) should be high—around 1. But the probability of the consequent, conditional on the antecedent, can't be high, because the consequent has (per TVT) low probability and the antecedent high.

Instead of going by way of conditional probabilities, your credence intuitively goes by the conditional probability of the consequent of the *imported* conditional, conditional on its antecedent:

(4) If a Republican wins and Reagan doesn't, then Anderson will.

So in general, probabilistic modus ponens for complex conditionals seems invalid, while probabilistic import-export seems valid:

10.2 Embedded conditionals

In McGee's case, (5) seems very probable.

(5) Either Carter will win, or else, if it's not Reagan, it will be Anderson.

But the probability of Carter is low; and the probability of Anderson, conditional on it not being Reagan, is extremely low. But, by probability theory, $P(a \vee b) \leq P(a) + P(b)$. So, per TVT, the probability of (5) should be small!

I think this is an underappreciated fact about McGee's case: it doesn't just show something about how The Thesis interacts with conditionals with complex consequents; it shows something more general about how conditionals, even simple ones, interact with their local contexts.

10.3 Interlude: How we capture these facts

Well, in the same way as in the logical cases.

If R can access all and only context-set worlds, then unembedded $p > q$ will have the probability of q conditional on p.

But if we limit the admissible vectors by embedding $p > q$ (and hence limiting the satt vectors), judgments will change, in line with the above observations.

10.4 Urn cases

The same is true if we restrict R. We can get counterexamples to TVT even for unembedded simple conditionals in cases where intuitively we hold fixed more than just the context.

Pollock 1981; McGee 2000; Kaufmann 2004; Rothschild 2013

> David is standing in front of two urns of marbles. He will flip a fair coin to choose an urn. If the coin lands heads, he will choose a marble at random from Urn 1; if tails, he will choose a marble at random from Urn 2. Each urn contains ten marbles. In Urn 1, there are eight red and two black marbles; all of them have a small white spot on them. In Urn 2, there are eight black and two red marbles; none of those is spotted.

In sum: Urn 1=10 marbles, all spotted, 8 red; Urn 2=10 marbles, no spotted, 2 red. Consider (6):

(6) If David picks a red marble, then it will be spotted.

What credence should you have in (6)?

- One answer: for any marble from Urn 1, if it's red, it's spotted. For any marble from Urn 2, if it's red, it's not spotted. So (6) is true just in case the marble is from Urn 1; but your credence in that is .5.

call this the *non-conforming* judgment, and the next judgment the *conforming* judgment.

- Another natural answer: there are ten red marbles total, which are chosen with equal probability. Of these, eight are spotted, and two are not spotted. So your credence in (6) should be .8.

So it seems like there are two natural perspectives on the credence of (6), only one of which yields the conditional probability .8.

Cases like this are easy to multiply.

10.5 Some replies

A skeptical response: following Kaufmann 2004, you might think this is just a failure to apply the law of total probability correctly. There are two salient options, Urn 1 and Urn 2. So people calculate:

$$P(S|R) =^{?} P(S|R, U1)P(U1) + P(S|R, U2)P(U2) = .5$$

But this is just a mistake; the correct calculation is instead:

$$P(S|R) = P(S|R, U1)P(U1|R) + P(S|R, U2)P(U2|R) = .8$$

But vary the case so that David first flips a coin to decide whether to proceed. If heads, he simply doesn't choose at all; otherwise, things proceed as above. But now there are three salient options: Urn 1, Urn 2, and neither; and now we cannot proceed by partition:

if we simply dropped the undefined term, we'd get .25, not the intuitive .5. If we renormalize we'd get .5—but this is an increasingly baroque error theory.

$$P(S|R, U1)P(U1) + P(S|R, U2)P(U2) + \underbrace{P(S|R, \overline{U1 \vee U2})}_{\#} P(\overline{U1 \vee U2}) =??$$

So I don't see a reason to think the non-conforming judgment is a mistake.

You might worry that the non-conforming reading will turn you into a money-pump. I don't think so. Consider:

Cf. Wójtowicz and Wójtowicz 2021.

> Ginger is standing before an urn containing 60 double-heads coins, 20 fair coins, and 20 double-tailed coins. She will choose a coin without looking, and then look at it and decide whether to flip it. Ginger is disinclined to flip double-headed coins: conditional on the coin being double-headed, the chance she'll flip it is only $\frac{1}{3}$. But if the coin is fair or double-tailed, she will certainly flip it. What's the probability of (7)?
>
> (7) If Ginger flips the coin she chooses, it will land heads.

- The conditional probability of the coin landing heads, conditional on being flipped, is .5.
- But another, equally prominent judgment can be brought out as follows. Of the equiprobable coins that Ginger could select, 60 out of 100 will certainly land heads, and 20 out of 100 will land heads with probability .5, and 20 certainly won't land heads. So (7) is certainly true of 60% of the coins, and has probability .5 of another 20%; and thus has total probability .7.

This case brings out the rationality of the non-conforming judgment, via betting behavior. Suppose that you are going to take a bet on (7); what odds are fair? One intuition you can have is that, if the selected coin is never flipped, the bet will be called off; if the coin is flipped and lands heads, you win; and if it's flipped and lands tails, the bookie wins. Then you should be willing to pay \$.50 for a bet that pays \$1 iff (7) is true, since you win the bet in exactly half the cases where it is not called off.

But if you took a bet on (7), then discovered that the selected coin was double-headed, you could equally insist you have won the bet—even if the coin is in fact never flipped. In that case, you can reasonably be certain that (7) is true. This seems like *a* rational way to call bets. But if you bet that way, then you should be willing to pay \$.70 for a bet that pays \$1 iff (7) is true, since you will win the bet (on this way of assessing it) in .7 of all possible (equiprobable) cases.

In general, when R is an equivalence relation, we have $P_c(A > C) = P_c(AC) + \sum_{w \in \neg A\kappa_c} P_c(w) \cdot P_c(C|A \cap R(w))$, given maximality. When R is not an equivalence relation, this won't necessarily hold.

11 Freedom and resentment

This builds on, but I think improves, vF's approach, in a way that speaks to some broader points that I think are important:

- the failure of TT in the case of complex conditionals is not due to the semantics of conditionals in particular, since we find an exactly parallel phenomenon for disjoined simple conditionals. It is, rather, due to the interaction of conditionals with their local contexts.
- in general, local contexts constrain interpretation—in a way that shows up not just in compatibility and inference judgments, but also in probability judgments
- but that constraint is not deterministic: interpretation is context-sensitive in ways that show up in the non-conforming judgments. Local contexts bound interpretation but do not determine it.

References

Adams, E. (1975). *The Logic of Conditionals*. Dordrecht.

Ellis, B. (1978). A unified theory of conditionals. *Journal of Philosophical Logic*, pages 107–124.

van Fraassen, B. (1976). Probabilities of conditionals. In Harper and Hooker, editors, *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, volume I, pages 261–308. D. Reidel Publishing Company, Dordrecht-Holland.

Kaufmann, S. (2004). Conditioning against the grain. *Journal of Philosophical Logic*, 33:583–606.

Lewis, D. (1976). Probabilities of conditionals and conditional probabilities. *The Philosophical Review*, 85(3):297–315.

McGee, V. (2000). To tell the truth about conditionals. *Analysis*, 60(1):107–111.

Pollock, J. L. (1981). Indicative conditionals and conditional probability. In Harper, W. L., Stalnaker, R., and Pearce, G., editors, *Ifs: Conditionals, Belief, Decision, Chance, and Time*, pages 249–252. Reidel.

Rothschild, D. (2013). Do indicative conditionals express propositions? *Noûs*, 47(1):49–68.

Stalnaker, R. (1974). Letter to Bas Van Fraassen.

Stalnaker, R. C. (1970). Probability and conditionals. *Philosophy of Science*, 37(1):64–80.

Wójtowicz, A. and Wójtowicz, K. (2021). Dutch book against Lewis. *Synthese*.