What we can learn from how trivalent conditionals avoid triviality

Daniel Lassiter
Stanford University

In press at *Inquiry*. Version of November 29, 2019.

Abstract A trivalent theory of indicative conditionals automatically enforces Stalnaker's thesis—the equation between probabilities of conditionals and conditional probabilities. This result holds because the trivalent semantics requires, for principled reasons, a modification of the ratio definition of conditional probability in order to accommodate the possibility of undefinedness. I explain how this modification is motivated and how it allows the trivalent semantics to avoid a number of well-known triviality results, in the process clarifying why these results hold for many bivalent theories. In short, the slew of triviality results published in the last 40-odd years need not be viewed as an argument against Stalnaker's thesis: it can be construed instead as an argument for abandoning the bivalent requirement that indicative conditionals somehow be assigned a truth-value in worlds in which their antecedents are false.

Keywords: Indicative conditionals, probability, trivalence

1 Lewis' original result

Stalnaker (1970) makes the intuitively reasonable proposal that the probability of the proposition that an English conditional denotes is systematically equal to the conditional probability of the consequent given the antecedent. Using '$A \Rightarrow C$' to pick out the proposition that the English indicative conditional 'If A then C' denotes (with appropriate substitutions for 'A' and 'C'):

(1) $$P(A \Rightarrow C) = P(C \mid A)$$

In addition to intuition, a large number of experimental studies support this equation: see for instance Evans & Over 2004; Douven & Verbrugge 2013 and references therein.

Problematically, Lewis (1976) proves that this equation leads to absurd consequences in the presence of several other apparently innocuous assumptions, namely:

i. The probabilities of conditionals behave like the probabilities of any other proposition: in particular, the law of total probability holds. For any B with $0 < P(B) < 1$,

$$\begin{aligned} P(A \Rightarrow C) &= P(A \Rightarrow C \wedge B) + P(A \Rightarrow C \wedge \neg B) \\ &= P(A \Rightarrow C \mid B) \times P(B) + P(A \Rightarrow C \mid \neg B) \times P(\neg B) \end{aligned}$$

ii. P is closed under conditionalization, so that—given equation (1)—$P(A \Rightarrow C \mid B) = P(C \mid A \wedge B)$ for any B.

Triviality follows immediately: taking $B = C$, (i) and (ii) give us

$$\begin{aligned} P(A \Rightarrow C) &= P(A \Rightarrow C \mid C) \times P(C) + P(A \Rightarrow C \mid \neg C) \times P(\neg C) \\ &= P(C \mid A \wedge C) \times P(C) + P(C \mid A \wedge \neg C) \times P(\neg C) \\ &= 1 \times P(C) + 0 \times P(\neg C) \\ &= P(C), \end{aligned}$$

as long as $0 < P(C) < 1$. So, a conditional's probability is not only equal to the probability of its consequent given its antecedent: if it is not 1 or 0, it is also equal to the probability of its consequent—clearly an unacceptable result.

Lewis' proof is not very enlightening as to *why* equation (1) leads to absurd results. In a footnote in a different paper, Lewis (1975) mentions obliquely that a trivalent semantics for the indicative conditional is able to avoid his triviality result, but asserts without explanation that trivalence is too high of a price ('exorbitant'). Whatever Lewis' reasons for thinking this, trivalent theories of the semantics of indicative conditionals have been a respectable though non-mainstream approach for a long time, going back at least to de Finetti (1936; English translation in de Finetti 1995) and continuing in philosophical logic through Jeffrey 1963; Cooper 1968; Belnap 1970; McDermott 1996; Milne 1997; Cantwell 2008; Huitink 2008; Rothschild 2014a; Égré & Cozic 2016 and a closely related proposal in Bradley 2002. There is also a substantial parallel literature in artificial intelligence (e.g., Benferhat, Dubois & Prade 1997) and a more recent literature in psychology (Politzer, Over & Baratgin 2010; Baratgin, Over & Politzer 2014; Baratgin, Politzer, Over & Takahashi 2018). Not only is this theory worth taking seriously, but thinking carefully about why Lewis' triviality result does not apply to it may help to clarify why the original triviality result does hold. I suggest that the culprit is the application of the standard ratio definition of conditional probability to conditionals, which the trivalent semantics requires us to modify for principled reasons. I then examine a number of further triviality proofs.[1] Several of these fail to go through in the trivalent system for exactly the same reason as Lewis', while others fail because they depend on reasoning about the probability of a bivalent conditional given the negation of its antecedent, which is always undefined in the trivalent system. In addition, in several cases the highly intuitively constraints on the probability/conditionals connection that motivate the triviality results turn out to be theorems of the trivalent system. This fact provides further reason to take trivalence seriously. Finally, I analyze the cardinality properties of bivalent and trivalent conditionals and explain how the differences make it possible to avoid Hájek's (1989) "Wallflower" result.

2 Trivalent semantics and probability

In de Finetti's (1936) trivalent truth-table, $A \Rightarrow C$ is

- true (1) if A and C are both true;
- false (0) if A is true and C is false;
- undefined (#) if A is false or undefined, or if C is undefined.

de Finetti motivates the undefinedness of a conditional with a false antecedent by considering bets on conditionals. For instance, imagine that Jim has accepted a bet on 'The Jets will win the game if they win the initial coin toss'. If the Jets do not win the coin toss, it would be perverse for the bookie to claim that

1 See Khoo & Santorio 2018 for an extremely valuable recent survey of triviality results.

Jim had lost the bet–or, for that matter, that he had won it. Instead, the bet should be simply called off. Experimental evidence indicates that naïve participants share this intuition (Politzer et al. 2010). On similar grounds, it seems arbitrary to assign either 'true' or 'false' to a conditional whose antecedent turns out false. In a trivalent system it is assigned neither value.

The trivalent semantics may make various predictions about nested conditionals and compounds of conditionals, depending on the definitions of other connectives and embedding operators. For the purposes of this paper I will assume only that the negation of # is #; and the conjunction of # with 1 or # is #. The Strong Kleene tables, for example, have these features. When dealing with right-nested conditionals, we will also make use of de Finetti's assumption that a conditional with an undefined antecedent is undefined.[2] For a discussion of alternative trivalent theories, including some that handle undefined antecedents differently, see Egré, Rossi & Sprenger 2019.

2.1 Trivalent probability: Two systems

Since the semantics is not bivalent, we have to modify the classical bivalent semantics for probability to cope with the # value. Cantwell (2006) provides the necessary modification: in effect, we leave everything alone except that the probability of a proposition A is normalised by the probability that A is defined. Intuitively, this means that we simply ignore worlds where the proposition is undefined when calculating probabilities. This strategy mirrors closely the way that trivalent theorists have addressed restricted quantification (Belnap 1970; McDermott 1996) and the use of conditionals to restrict quantifiers and sentential operators (Huitink 2008: §5). In each case, restriction to antecedent-satisfying scenarios is achieved by defining the relevant operators so that they do not consider individuals or worlds where the relevant conditional is undefined.

Extending this strategy to probability measures, consider a propositional language $\mathcal{L}$ that is closed under negation, disjunction, the trivalent conditional defined above, and a unary connective *True*. We allow that some sentences are undefined at some assignments, i.e., neither $True(A)$ nor $True(\neg A)$ holds. A unary operator TV ('has a truth-value') is defined as $TV(A) = True(A) \vee True(\neg A)$. A bivalent sentence is one that has a truth-value at every assignment. Cantwell (2006, 2008) axiomatises a non-bivalent probability measure $P_T : \mathcal{L} \to [0,1]$ on sentences of $\mathcal{L}$ as:

- For bivalent A and B of $\mathcal{L}$:
  - $P_T(A) = P_T(B)$ if A and B are logically equivalent.
  - $P_T(\neg A) = 1 - P_T(A)$.
  - $P_T(A \vee B) = P_T(A) + P_T(B)$ if A and B are inconsistent.
- For any A:
  - $P_T(A) = \frac{P_T(True(A))}{P_T(TV(A))}$, as long as $P_T(TV(A)) > 0$. (REST-E)

(REST-E is mnemonic for 'extensional restriction'.) For the bivalent (conditional-free) fragment of the language trivalent probability behaves exactly like ordinary probability. Cantwell (2006) proves that that an

2 Things are, of course, not really this simple. See Bradley 2002; Edgington 2014; Douven 2016 for criticisms of the trivalent semantics based on compounds and nestings of conditionals. These critiques depend on assumptions about how English *and* and *or* relate to $\wedge$ and $\vee$ so defined—and more generally how they handle undefinedness—which are disputed by McDermott (1996). That said, trivalent theories do face substantial challenges from embedded and nested conditionals which still need to be addressed in full detail.

agent who follows this system of non-bivalent probability is uniquely able to avoid Dutch Books for both unconditional bets and conditional bets.

Since we will be discussing both bivalent and trivalent probability, I will distinguish them by using 'P_B' for standard (Kolmogorovian) probability and P_T for trivalent probability.

Cantwell's non-bivalent probability differs from the familiar Kolmogorov 1933 definition in that it is not intensional—there is no mention of possible worlds. Since possible worlds play an important role in the semantics of conditionals and in certain of the triviality arguments discussed below, it is worth pausing to frame a intensional variant of Cantwell's definition. The intensional variant is close to the one given by Rothschild (2014a) (who apparently was unaware of Cantwell's work).

Rather than associating sentences of $\mathcal{L}$ directly with probabilities, we will now associate them with trivalent propositions which are then assigned probabilities in a unified format. Let W be a set of possible worlds. Specifying the trivalent proposition associated with a sentence A requires two sets of worlds: a domain $TV(A) \subseteq W$ of worlds in which the proposition is defined, and a truth-set $True(A) \subseteq TV(A)$ in which the proposition is defined and true. TV and $True$ are simply the intensional counterparts of the corresponding functions in Cantwell's definition. So, the interpretation function $[\![\cdot]\!]$ maps each sentence A to a pair of sets $\langle TV(A), True(A)\rangle$.

$[\![A]\!] = \langle TV(A), True(A)\rangle$ is

- # at w if $w \notin TV(A)$;
- true at w if $w \in True(A)$;
- false at w if $w \in TV(A) - True(A)$.

A bivalent proposition is one whose domain is W. A (non-trivially) trivalent proposition is one whose domain is a proper subset of W. For bivalent A and C, the trivalent conditional is given by

$$[\![A \Rightarrow C]\!] = \langle True(A), True(A) \cap True(C)\rangle.$$

For possibly trivalent A and C, the following refinement ensures that a conditional with an undefined antecedent or consequent is undefined:

$$[\![A \Rightarrow C]\!] = \langle True(A) \cap TV(C), True(A) \cap True(C)\rangle.$$

The intensional variants of the $\wedge$ and $\neg$ defined above are:

- $[\![A \wedge B]\!] = \langle TV(A) \cap TV(B), True(A) \cap True(B)\rangle$
- $[\![\neg A]\!] = \langle TV(A), TV(A) - True(A)\rangle$

We can now state a simple intensional definition of the trivalent probability measure $P_T : \mathcal{P}(W)^2 \to [0,1]$, which takes as its argument trivalent propositions (i.e., pairs of classical propositions; for readability I will write $P_T(D,E)$ instead of $P_T(\langle D,E\rangle)$). For any subsets D and E of W:

- $P_T(W,W) = 1$;
- $P_T(W, D \cup E) = P_T(W,D) + P_T(W,E)$ if $D \cap E = \varnothing$;
- $P_T(D,E) = \frac{P_T(W,D \cap E)}{P_T(W,D)}$. (REST-I)

The first and second axioms correspond exactly to standard bivalent axioms; and the crucial third axiom requires that the probability of any proposition is computed relative to the domain in which it is defined. This axiom is the intensional counterpart of Cantwell's REST-E. Since a bivalent proposition $\langle W, D\rangle$ is defined everywhere, the third axiom places on bivalent propositions only the trivial requirement that $P_T(W, E) = P_T(W, W \cap E)/P_T(W, W) = P_T(W, E)/1$. It is a straighforward exercise to show that in the intensional system the composition of $[\![\cdot]\!]$ with P_T enforces Cantwell's axioms.

For example, consider the model that Hájek (1989) uses to illustrate his "Wallflower" result for bivalent probability (discussed in more detail in section 12 below). In this model, there are three possible worlds a, b, and c, each of which has probability $1/3$. The full Boolean algebra is given on the left of Figure 1, with propositions labeled with their probabilities. On the right side we have the subalgebra based on $\{b, c\}$, which is the region in which conditionals with antecedent "If $\{b, c\}$" concentrate all of their probability. Note that the probability of each of these conditionals is the probability of the consequent normalized by dividing by the probability of the antecedent—$P_T(\{b, c\}) = \frac{2}{3}$—in accordance with the third axiom of trivalent probability (REST-I).

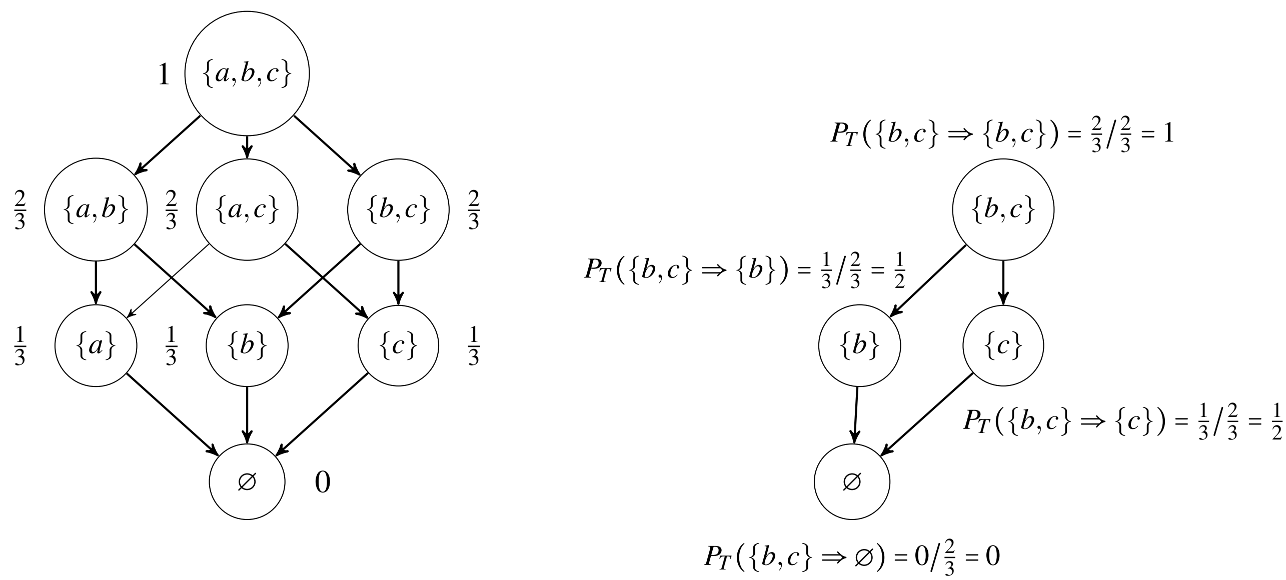

Figure 1 Left: Boolean algebra based on three worlds. Right: Subalgebra based on the worlds in which trivalent conditional of the form "If $\{b, c\}$ then A" concentrate their probability.

Note that this representation does not imply that these are the only conditionals with antecedent $\{b, c\}$ that are nontrivial; rather, any nontrivial conditional with antecedent $\{b, c\}$ is logically equivalent to a conditional with the same antecedent whose consequent lies in the subalgebra on the right side of Figure 1. This is because $A \Rightarrow C$ is equivalent to $A \Rightarrow (A \wedge C)$: we can restrict the consequent by the antecedent without loss of information. For example, $\{b, c\} \Rightarrow \{a, b\}$ is a nontrivial conditional with antecedent $\{b, c\}$ whose consequent does not lie in the subalgebra on the right of Figure 1. However, it is equivalent to $\{b, c\} \Rightarrow \{b\}$, whose consequent does lie in this subalgebra. Both conditionals are undefined in world a, true in b, and false in c.

3 Stalnaker's thesis

Trivalent semantics and probability immediately enforce Stalnaker's thesis for bivalent antecedents and consequents. Assuming $P_T(A) > 0$,

$$\begin{aligned} P_T(A \Rightarrow C) &= \frac{P_T(True(A \Rightarrow C))}{P_T(TV(A \Rightarrow C))} \\ &= \frac{P_T(A \wedge C)}{P_T(A)} \\ &= P_T(C \mid A). \end{aligned}$$

This result is also illustrated for the intensional variant in Figure 1. In the trivalent system this restricted form of Stalnaker's thesis is not an empirical hypothesis, but rather a theorem that results from the fact that the probability of every sentence is normalised by the probability that the sentence is defined. Therefore the thesis places no additional constraints on probability distributions beyond what follows from the definitions of the connectives and of trivalent probability.

Assuming that a conditional with an undefined antecedent is undefined, Stalnaker's thesis also holds for right-nested conditionals (with bivalent literals):

$$P_T(A \Rightarrow (B \Rightarrow C)) = P_T(B \Rightarrow C \mid A) = P_T(C \mid A \wedge B) = \frac{P_T(A \wedge B \wedge C)}{P_T(A \wedge B)}$$

See sections 9 and 10 below for more discussion.

4 Conditional probabilities in the trivalent setting

This section draws out the modification of the ratio definition of conditional probability that is implied by Cantwell's (2006; 2008) discussion, though he does not discuss the issue explicitly. I then describe some crucial differences from standard probability that emerge from this modification.

The familiar ratio formula can remain unchanged for bivalent C and A—

$$P_T(C \mid A) = \frac{P_T(C \wedge A)}{P_T(A)}, \text{ provided } P_T(A) > 0$$

—but this similarity hides a difference from standard probability that will be important in what follows. As noted above, we assume that the conjunction of 1 or # with # is #. Therefore, if A or C is not bivalent, the calculation of $P_T(C \wedge A)$ depends on the probability that each has a truth-value. From REST-E we have

$$P_T(C \wedge A) = \frac{P_T(True(C \wedge A))}{P_T(TV(C \wedge A))},$$

assuming $P_T(TV(C \wedge A)) > 0$. As a result the possibility of undefinedness also affects the ratio formula:

$$P_T(C \mid A) = \frac{P_T(True(C \wedge A))}{P_T(TV(C \wedge A) \wedge True(A))},$$

provided that $P_T(TV(C \wedge A) \wedge True(A)) > 0$. The normalising constant is the probability that the conjunction is defined *and* the condition A is true.

If C happens to be a conditional $D \Rightarrow E$—as it will in a number of the triviality proofs examined below—the ratio formula implies (for bivalent D, E, A)

$$P_T(D \Rightarrow E \mid A) = \frac{P_T(True(D \Rightarrow E \wedge A))}{P_T(TV(D \Rightarrow E \wedge A) \wedge True(A))} \tag{2}$$

$$= \frac{P_T(D \wedge E \wedge A)}{P_T(D \wedge A)}. \tag{3}$$

The conditional probability of a trivalent conditional thus depends on normalisation by the probability of the *conjunction* of the conditioning proposition and the conditional's antecedent.

This normalisation is the key to understanding why the trivalent system avoids several well-known triviality proofs: a key equivalence that appears in these arguments holds only for bivalent propositions. In bivalent probability, the following equation holds unrestrictedly provided $P_B(B) > 0$:

$$P_B(D \wedge B) = P_B(D \mid B) \times P_B(B)$$

This is of course because of the ratio definition of conditional probability:

$$\begin{aligned} P_B(D \mid B) \times P_B(B) &= \frac{P_B(D \wedge B)}{P_B(B)} \times P_B(B) \\ &= P_B(D \wedge B) \end{aligned}$$

This equivalence does *not* hold for P_T when D is non-trivially trivalent. Instead we have (for bivalent A, C, B):

$$P_T(A \Rightarrow C \mid B) \times P_T(B) = \frac{P_T(A \wedge C \wedge B)}{P_T(A \wedge B)} \times P_T(B)$$

which does not reduce to $P_T(A \Rightarrow C \wedge B)$ because of the nonstandard normalisation. This difference from standard probability will play a starring role in our discussion of Lewis's (1976) triviality result and several others. In bivalent probability, we would have instead the equivalence

$$P_B(A \Rightarrow C \wedge B) = P_B(A \Rightarrow C \mid B) \times P_B(B). \tag{4}$$

Formulas instantiating (4) will appear repeatedly in the triviality proofs examined below, and replacing it with (3) allows us to block the proofs in a motivated way.

5 Lewis's (1976) proof

The pivotal moment in Lewis's (1976) proof is the transition from line (5) to (6).

$$P_B(A \Rightarrow C) = P_B(A \Rightarrow C \wedge C) + P_B(A \Rightarrow C \wedge \neg C) \tag{5}$$

$$= P_B(A \Rightarrow C \mid C) \times P_B(C) + P_B(A \Rightarrow C \mid \neg C) \times P_B(\neg C) \tag{6}$$

$$= P_B(C \mid A \wedge C) \times P_B(C) + P_B(C \mid A \wedge \neg C) \times P_B(\neg C) \tag{7}$$

$$= 1 \times P_B(C) + 0 \times P_B(\neg C) \tag{8}$$

In motivating steps (5) and (6) above, I rationalised them by noting the sensible desideratum that 'probabilities of conditionals behave like the probabilities of any other proposition'. But there are principled reasons in the trivalent semantics **not** to treat the probabilities of conditionals in the same way that we treat those of bivalent sentences, particularly where the ratio definition of conditional probability is concerned.

As described above, the transition from (5) to (6) is not legitimate in trivalent probability: $P_T(A \Rightarrow C \wedge C)$ does not in general equal $P_T(A \Rightarrow C \mid C) \times P_T(C)$. Instead we have (for bivalent A, C, and assuming $P_T(A \wedge C), P_T(A \wedge \neg C) > 0$):

$$\begin{aligned} P_T(A \Rightarrow C) &= P_T(A \Rightarrow C \wedge C) + P_T(A \Rightarrow C \wedge \neg C) \\ &= \frac{P_T(A \wedge C \wedge C)}{P_T(A)} + \frac{P_T(A \wedge C \wedge \neg C)}{P_T(A)} \\ &= \frac{P_T(A \wedge C)}{P_T(A)} + 0 \end{aligned}$$

which does not have any interesting relation to the trivializing term $P_T(A \Rightarrow C \mid C)$ as it would in bivalent semantics. Indeed, this derivation is simply a roundabout way to get to Stalnaker's thesis. In sum, nothing of interest follows in the trivalent setting from the application of the law of total probability to conditionals.[3]

Lewis' proof has been taken to show that probability and conditionals do not get along well. I suggest that this appearance holds because we have been trying to apply to conditionals rules of probability that are only appropriate where bivalent sentences are concerned. This diagnosis is also appropriate for a number of the other triviality proofs to be considered in the rest of the paper.

6 Milne 2003

Using a proof very similar to Lewis', Milne (2003) proves that the probability of a conditional $A \Rightarrow C$ is equal to the probability of the material conditional $A \supset C$. Given this, Stalnaker's thesis can hold only for trivial bivalent probability models, since $P_B(C \mid A) = P_B(A \supset C)$ only if $P_B(A) = 1$, $P_B(C \mid A) = 1$, or $P_B(A \wedge C) = 0$.

The trivalent semantics avoids this triviality result for the same reason that it avoids Lewis': the crucial moment in the proof has

(9) $$P_B(A \Rightarrow C) = P_B(A \Rightarrow C \wedge A \supset C) + P_B(A \Rightarrow C \wedge \neg A \supset C)$$

(10) $$\geq P_B(A \Rightarrow C \wedge A \supset C)$$

(11) $$= P_B(A \Rightarrow C \mid A \supset C) \times P_B(A \supset C)$$

(12) $$= P_B(A \supset C),$$

assuming $P_B(A \wedge C) > 0$.

The transition from (10) to (11) is not legitimate in the trivalent semantics. Instead, for bivalent A with $P_T(A) > 0$,

$$\begin{aligned} P_T(A \Rightarrow C \wedge A \supset C) &= \frac{P_T(A \wedge C \wedge A \supset C)}{P_T(A)} \\ &= \frac{P_T(A \wedge C)}{P_T(A)} \\ &= P_T(A \Rightarrow C), \end{aligned}$$

3 Note that this diagnosis differs from that of Cantwell (2008: 174), who states without elaboration that Lewis' result is avoided because trivalent probabilities can be non-additive. The proof does rely on additivity in its application of the law of total probability, but this particular instance appears to be legitimate.

which is as it should be. (Whether $A \Rightarrow C$ *entails* $A \supset C$ depends on what is the right concept of consequence for trivalent semantics, which is a matter of active investigation: see Chemla, Égré & Spector 2017; Chemla & Egré 2019; Egré et al. 2019. This entailment does hold, for example, for a notion of consequence $\vDash_T$ where $A \vDash_T B$ just in case there is no assignment in which A is not false and B is false.)

7 Bradley 2007

Bradley (2007) provides an alternative proof of triviality which does not depend on Stalnaker's thesis but instead on two very plausible assumptions: for any rational P',

- $P'(C) = 1$ then $P'(A \Rightarrow C) = 1$.
- If $P'(C) = 0$ then $P'(A \Rightarrow C) = 0$.

Following Lewis' proof through line (6) and then using these these two assumptions we can trivialise P_B without relying on Stalnaker's thesis, which—together closure under conditionalization—licenses the transition from (6) to (7). Now we can get directly from (14) to (8)/(15), simply choosing $P' = P_B(\cdot \mid C)$ in the first instance and $P' = P_B(\cdot \mid \neg C)$ in the second.

$$P_B(A \Rightarrow C) = P_B(A \Rightarrow C \wedge C) + P_B(A \Rightarrow C \wedge \neg C) \tag{13}$$

$$= P_B(A \Rightarrow C \mid C) \times P_B(C) + P_B(A \Rightarrow C \mid \neg C) \times P_B(\neg C) \tag{14}$$

$$= 1 \times P_B(C) + 0 \times P_B(\neg C) \tag{15}$$

In the trivalent setting, Bradley's proof fails for the same reason that it does in Lewis' and Milne's: the transition from (13) to (14) is not legitimate. Reassuringly, Bradley's very plausible assumptions are also theorems of the trivalent system, at least for bivalent A and C with $P_T(A) > 0$.

- If $P_T(C) = 1$ then $P_T(A \Rightarrow C) = P_T(A \wedge C)/P_T(A) = P_T(A)/P_T(A) = 1$.
- If $P_T(C) = 0$ then $P_T(A \Rightarrow C) = P_T(A \wedge C)/P_T(A) = 0/P_T(A) = 0$.

8 Bradley 2000

Bradley (2000) shows that a certain highly intuitive 'Preservation Condition' cannot hold. The condition is that for any A, C such that $A \nvDash C$ and $A \Rightarrow C \nvDash C$,

$$\text{If } P_B(A) > 0 \text{ and } P_B(C) = 0\text{, then } P_B(A \Rightarrow C) = 0.$$

The proof depends on reasoning about the probability of the conjunction $A \Rightarrow C \wedge \neg A$, which is non-trivial in bivalent systems but trivially 0 or undefined in the trivalent system. This is because all possible worlds in which $\neg A$ holds are worlds in which $A \Rightarrow C$ has the value #, and so the conjunction is never true.

$$P_T(A \Rightarrow C \wedge \neg A) = \frac{P_T(A \wedge C \wedge \neg A)}{P_T(A)} = \frac{0}{P_T(A)}$$

In the trivalent system, if $P_T(A) > 0$ and $P_T(C) = 0$ we have (since Stalnaker's thesis holds)

$$P_T(A \Rightarrow C) = \frac{P_T(A \wedge C)}{P_T(A)} = \frac{0}{P_T(A)},$$

with the result that the trivalent variant of Bradley's Preservation Condition is a theorem.

9 Fitelson 2015

The principle of Import-Export is valid in a trivalent semantics, as long as a conditional with an undefined antecedent is undefined.

$$A \Rightarrow (B \Rightarrow C) \equiv (A \wedge B) \Rightarrow C$$

Both sides of the equivalence will be true if $A \wedge B \wedge C$, false if $A \wedge B \wedge \neg C$, and undefined otherwise.

Since these conditionals are logically equivalent, the principle obviously places no additional constraint on probability distributions beyond what follows from the semantics and the definition of probability. In contrast, in the bivalent setting both Import-Export and Stalnaker's thesis are substantive assumptions, and—as Fitelson (2015) shows—can be used to derive what Fitelson calls the 'Resilient Equation': for any Y,

$$P_B(A \Rightarrow C \mid Y) = P_B(C \mid A \wedge Y)$$

This equation holds also in the trivalent setting (at least, for bivalent A, C, Y), not as a substantive hypothesis but as a consequence of the fact that trivalent conditional probabilities are normalised also by the probability that the propositions involved are defined (cf. Equation (3) in section 4 above).

$$P_T(A \Rightarrow C \mid Y) = \frac{P_T(True(A \Rightarrow C \wedge Y))}{P_T(TV(A \Rightarrow C \wedge Y) \wedge True(Y))}$$

For bivalent A, C, Y, this is equivalent to

$$P_T(A \Rightarrow C \mid Y) = \frac{P_T(A \wedge C \wedge Y)}{P_T(A \wedge Y)}$$
$$= P_T(C \mid A \wedge Y)$$

Since the Resilient Equation is simply a theorem, it clearly does not have the untoward consequences in the trivalent setting that it does in the bivalent setting; still, it may be useful to go through the main steps of Fitelson's (2015) result to see where the differences lie.

Fitelson works through three instances of the Resilient Equation and shows that their conjunction leads to triviality. To preview: the first two instances make crucial use of the ratio definition that appears differently in the trivalent system, and the third instance is Stalnaker's Thesis itself, which we have seen is unproblematic in trivalent semantics.

- The $\neg C$-instance: $P_B(A \Rightarrow C \mid \neg C) = P_B(C \mid A \wedge \neg C)$, and so (by the bivalent ratio definition)

  $$\frac{P_B((A \Rightarrow C) \wedge \neg C)}{P_B(\neg C)} = \frac{P_B(C \wedge A \wedge \neg C)}{P_B(\neg C)} = 0,$$

  which implies that $P_B(A \Rightarrow C \wedge \neg C) = 0$.
- The $A \supset C$-instance: $P_B(A \Rightarrow C \mid A \supset C) = P_B(C \mid A \wedge A \supset C)$, and so (by the bivalent ratio definition)

  $$\frac{P_B(A \Rightarrow C \wedge A \supset C)}{P_B(A \supset C)} = P_B(C \mid A \wedge C) = 1.$$

- The tautology: if we substitute a tautology for Y we get Stalnaker's Thesis: $P_B(A \Rightarrow C) = P_B(C \mid A)$.

By analyzing the constraints placed on a bivalent truth-table by these equations, Fitelson shows that the resulting distribution places positive probability only on two lines of the table: the one where A, C and $A \Rightarrow C$ are all true, and the line where A is true and C and $A \Rightarrow C$ are false. (By a striking coincidence (?), these two lines describe precisely the truth- and falsity-conditions of the trivalent conditional.) So, the resilient equation can hold in bivalent semantics only in trivial models in which the antecedent of a conditional is always true, and the conditional is true iff its consequent is.

In contrast, in the trivalent setting we have (provided all the conditional probabilities are defined, and for bivalent A, C)

- The $\neg C$-instance: $P_T(A \Rightarrow C \mid \neg C) = P_T(C \mid A \wedge \neg C)$, and so (by the trivalent ratio definition)

$$\frac{P_T(A \wedge C \wedge \neg C)}{P_T(A \wedge \neg C)} = \frac{P_T(C \wedge A \wedge \neg C)}{P_T(A \wedge \neg C)},$$

  which is trivially true.

- The $A \supset C$-instance: $P_T(A \Rightarrow C \mid A \supset C) = P_T(C \mid A \wedge A \supset C)$, and so (by the trivalent ratio definition)

$$\frac{P_T(A \wedge C \wedge A \supset C)}{P_T(A \wedge A \supset C)} = \frac{P_T(C \wedge A \wedge A \supset C)}{P_T(A \wedge A \supset C)},$$

  which is also trivially true.

- The tautology: if we substitute a tautology for Y we get Stalnaker's Thesis: $P_T(A \Rightarrow C) = P_T(C \mid A)$, which is a theorem as we saw above.

In sum, the trivalent semantics trivialises Fitelson's proof by converting the key moves in the proof, and the constraints that they produce, into obvious equalities enforced by the modified ratio definition.

10 Stalnaker's thesis for right-nested conditionals

The reasoning in the previous section assumed bivalence except where a sentence was explicitly marked as a conditional. One important question is whether the trivalent semantics can avoid another path to triviality via right-nested conditionals.[4] As we saw above, Stalnaker's thesis can (and should!) be applied to conditionals with conditional consequents. Using P_X as schematic for either P_B or P_C:

$$P_X(A \Rightarrow (B \Rightarrow C)) = P_X(B \Rightarrow C \mid A).$$

It follows from the Resilient Equation that

$$\begin{aligned} P_X(C \Rightarrow (B \Rightarrow C)) &= P_X(B \Rightarrow C \mid C) \\ &= P_X(C \mid B \wedge C) \\ &= 1. \end{aligned}$$

By similar reasoning we have $P_X(\neg C \Rightarrow (B \Rightarrow C)) = P_X(B \Rightarrow C \mid \neg C) = 0$.

4 Posed to me by a helpful reviewer, along with a sketch of the alternate triviality proof described here.

In the bivalent system, we can taking these results as inputs to the bivalent ratio formula, ending up at the last steps of Lewis' proof:

$$\begin{aligned} P_B(B \Rightarrow C) &= P_B(B \Rightarrow C \wedge C) + P_B(B \Rightarrow C \wedge \neg C) \\ &= P_B(B \Rightarrow C \mid C) \times P_B(C) + P_B(B \Rightarrow C \mid \neg C) \times P_B(\neg C) \\ &= 1 \times P_B(C) + 0 \times P_B(\neg C) \\ &= P_B(C), \end{aligned}$$

provided $P_B(C), P_B(\neg C) > 0$. This result could, then, be construed as an argument against extending Stalnaker's thesis to right-nested conditionals despite the intuitive plausibility of the extension.

In the trivalent system we do not have a choice about extending Stalnaker's thesis to right-nested conditionals: for bivalent A, B, C,

$$\begin{aligned} P_T(A \Rightarrow (B \Rightarrow C)) &= P_T(B \Rightarrow C \mid A) \\ &= P_T(C \mid A \wedge B) \\ &= \frac{P_T(A \wedge B \wedge C)}{P_T(A \wedge B)}, \end{aligned}$$

with the result that $P_T(C \Rightarrow (B \Rightarrow C)) = P_T(B \Rightarrow C \mid C) = 1$ (if defined). By analogous reasoning, $P_T(\neg C \Rightarrow (B \Rightarrow C)) = P_T(B \Rightarrow C \mid \neg C) = 0$ (if defined). This does not, however, provide an alternate route to Lewis-style triviality. Instead, we have only

$$\begin{aligned} P_T(B \Rightarrow C) &= P_T(B \Rightarrow C \wedge C) + P_T(B \Rightarrow C \wedge \neg C) \\ &= \frac{P_T(B \wedge C)}{P_T(B)} + \frac{P_T(B \wedge C \wedge \neg C)}{P_T(B)} \\ &= P_T(C \mid B) + 0, \end{aligned}$$

provided $P_T(C), P_T(\neg C) > 0$—just as in the discussion of Lewis' original proof. In the attempt to replicate the trivality proof in the trivalent setting, the key difference is again the nonstandard normalisation: there is no special relationship between the term $P_T(B \Rightarrow C \wedge C)$ and the trivializing term $P_T(B \Rightarrow C \mid C)$. In the first case, we have the decidedly non-trivializing

$$P_T(B \Rightarrow C \wedge C) = \frac{P_T(B \wedge C)}{P_T(B)} = P_T(C \mid B),$$

while the trivalizing term that is guaranteed to be equal to 1 expands to the reassuringly trivial-looking

$$P_T(B \Rightarrow C \mid C) = \frac{P_T(B \wedge C)}{P_T(B \wedge C)}.$$

This discussion also clarifies an argument made by Khoo & Santorio (2018) against the extension of Stalnaker's thesis to right-nested conditionals. Khoo & Santorio note that, if we assume the very plausible principle of Weak Centering—that $A \Rightarrow B$ entails $A \supset B$—we can derive bizarre probabilistic predictions from Stalnaker's thesis as applied to right-nested conditionals. Consider a roll of a fair die.

(16) If the die landed even, then, if it didn't land on 2 or 4, it landed on 6.

As they note, (16) should clearly receive probability 1. What about the antecedent and consequent in isolation?

(17) The die landed even.

(17) clearly has probability 1/2.

(18) If the die didn't land on 2 or 4, it landed on 6.

Presumably (18) has probability $1/4$, since the antecedent leaves 4 equiprobable options open. Summarizing, we have the intuitive probability assignments

$$\begin{aligned}
&P_B(\mathbf{even}) = 1/2\\
&P_B(\mathbf{even} \Rightarrow [\neg(\mathbf{two} \vee \mathbf{four}) \Rightarrow \mathbf{six}]) = 1\\
&P_B(\neg(\mathbf{two} \vee \mathbf{four}) \Rightarrow \mathbf{six}) = 1/4
\end{aligned}$$

But wait: by Weak Centering, (18) entails that the die came up even. As a result, the conjunction of (17) and (18) has the same probability as (18):

$$P_B([\neg(\mathbf{two} \vee \mathbf{four}) \Rightarrow \mathbf{six}] \wedge \mathbf{even}) = P_B(\neg(\mathbf{two} \vee \mathbf{four}) \Rightarrow \mathbf{six}) = 1/3$$

Now, by the extension of Stalnaker's thesis to right-nested conditionals we have

$$P_B(\mathbf{even} \Rightarrow [\neg(\mathbf{two} \vee \mathbf{four}) \Rightarrow \mathbf{six}]) = P_B(\neg(\mathbf{two} \vee \mathbf{four}) \Rightarrow \mathbf{six} \mid \mathbf{even}).$$

In the bivalent system that Khoo and Santorio assume, we can expand this equation via the bivalent ratio formula to derive

$$\begin{aligned}
P_B(\mathbf{even} \Rightarrow [\neg(\mathbf{two} \vee \mathbf{four}) \Rightarrow \mathbf{six}]) &= P_B(\neg(\mathbf{two} \vee \mathbf{four}) \Rightarrow \mathbf{six} \mid \mathbf{even})\\
&= \frac{P_B(\neg(\mathbf{two} \vee \mathbf{four}) \Rightarrow \mathbf{six} \wedge \mathbf{even})}{P_B(\mathbf{even})}\\
&= \frac{P_B(\neg(\mathbf{two} \vee \mathbf{four}) \Rightarrow \mathbf{six})}{P_B(\mathbf{even})}\\
&= \frac{1/4}{1/2}\\
&= 1/2.
\end{aligned}$$

This contradicts the intuition that (16) has probability 1.

If we assume bivalence, this argument provides support for Khoo & Santorio's conclusion that Stalnaker's Thesis is not correct for right-nested conditionals. However, as we have seen, the extension to right-nested conditionals is both intuitively plausible and a consequence of the trivalent system that we are considering. This is not a problem because Khoo & Santorio's derivation depends crucially on an application of the bivalent ratio formula that is inappropriate in the trivalent context. In the latter, we also have to normalise the term in the second line of the last derivation by the probability that the antecedent of the conditional is defined **and** the die is even, yielding

$$\begin{aligned} P_T(\mathbf{even} \Rightarrow [\neg(\mathbf{two} \vee \mathbf{four}) \Rightarrow \mathbf{six}]) &= P_T(\neg(\mathbf{two} \vee \mathbf{four}) \Rightarrow \mathbf{six} \mid \mathbf{even}) \\ &= \frac{P_T(\neg(\mathbf{two} \vee \mathbf{four}) \Rightarrow \mathbf{six} \wedge \mathbf{even})}{P_T(\neg(\mathbf{two} \vee \mathbf{four}) \wedge \mathbf{even})} \\ &= \frac{P_T(\mathbf{six})}{P_T(\neg(\mathbf{two} \vee \mathbf{four}) \wedge \mathbf{even})}. \end{aligned}$$

Since the roll is both even and (neither two nor four) only if it is six, this is equivalent to

$$\frac{P_T(\mathbf{six})}{P_T(\mathbf{six})} = 1.$$

This consequence agrees with the intuitions that Khoo & Santorio report, while also conforming to Weak Centering and the extension of Stalnaker's thesis to right-nested conditionals. This result provides another piece of support for the trivalent system, since Stalnaker's thesis is no less plausible for right-nested conditionals than for simple conditionals. The argument against the extension is due to features of the bivalent system that we have independent reasons to abandon in favor of a trivalent system that conforms to our intuitive judgments about the probabilities of conditionals.

11 Stalnaker 1976

Stalnaker (1976) gives a triviality proof that relies on four assumptions, the first one weaker than Stalnaker's thesis as framed above.

- For some P_B, for any A, B, C: $P_B(A \Rightarrow C) = P_B(C \mid A)$.
- Strong centering: $A \supset [(A \Rightarrow C) \equiv C]$ is a tautology.
- $A \Rightarrow C$ is inconsistent with $\neg(A \Rightarrow C)$.
- $A \Rightarrow B, B \Rightarrow A$, and $A \Rightarrow C$ jointly imply $B \Rightarrow C$.

Stalnaker shows that these conditions are not jointly satisfiable. No P_B that verifies Stalnaker's thesis can satisfy the second, third, and fourth conditions. The proof relies crucially on bivalence: specifically, the requirement that $A \Rightarrow C$ can true at worlds where A is false, so that it is possible for the conjunction of a conditional with the negation of its antecedent to have positive probability. Stalnaker's proof shows that the first assumption (weak Stalnaker's thesis) requires that $P_B(A \Rightarrow C \mid \neg C)$ is equal to $P_B(C \mid A)$. This consequence is essentially due to bivalence: conditionals obviously do not entail the negation of their antecedents, and so (in the bivalent setting) the probability of a conditional must somehow be distributed among the worlds where its antecedent is false. Using the law of total probability and strong centering we get

$$\begin{aligned} P_B(A \Rightarrow C) &= P_B(A \Rightarrow C \wedge A) + P_B(A \Rightarrow C \wedge \neg A) \\ &= P_B(A \wedge C) + P_B(A \Rightarrow C \wedge \neg A) \end{aligned}$$

Substituting $P_B(C \mid A)$ for $P_B(A \Rightarrow C)$ on the left side per Stalnaker's thesis we now have

$$P_B(C \mid A) = P_B(A \wedge C) + P_B(A \Rightarrow C \wedge \neg A)$$

and so

$$\begin{aligned}
P_B(A \Rightarrow C \wedge \neg A) &= P_B(C \mid A) - P_B(A \wedge C) \\
&= P_B(C \mid A) - \frac{P_B(A \wedge C)}{P_B(A)} \times P_B(A) \\
&= P_B(C \mid A) - P_B(C \mid A) \times P_B(A) \\
&= P_B(C \mid A) \times (1 - P_B(A)) \\
&= P_B(C \mid A) \times P_B(\neg A)
\end{aligned}$$

(From here the consequence that $P_B(A \Rightarrow C \mid \neg A) = P_B(C \mid A)$ follows by application of the bivalent ratio formula.) As a result, bivalence together with the first two assumptions of Stalnaker's proof implies that the conjunction of a conditional and the negation of its antecedent must have non-zero probability except in two trivializing cases, where either $P_B(C \mid A)$ or $P_B(\neg A)$ is equal to 0.

A trivalent system in fact does enforce all four of Stalnaker's assumptions (with an asterisk for the way that we spell out 'jointly imply' in the fourth condition for a trivalent system). But—as we saw already in the discussion of Bradley 2007 above—the conditional $A \Rightarrow C$ cannot have nonzero probability conditional on $\neg A$.

$$P_T(A \Rightarrow C \mid \neg A) = \frac{P_T(A \wedge C \wedge \neg A)}{P_T(A \wedge \neg A)} = \frac{0}{0}$$

12 Hájek's (1989) Wallflower result

Consider the model that Hájek (1989) uses to illustrate his 'Wallflower' result: a finite probability distribution on $W = \{a, b, c\}$, with each world receiving probability $1/3$. As Hájek notes, every non-zero unconditional probability in this model is a multiple of $1/3$, but some conditional probabilities are not. For instance, let A be a proposition true exactly at worlds a and b, and let C be true only at a. Then $P_B(C \mid A) = 1/2$. This means that the probability of the proposition denoted by the conditional $A \Rightarrow C$ cannot be equal to *any* conditional probability in this model: there are simply not enough unconditional probabilities to go around. This result generalises to any finite probability distribution.

The Wallflower result is essentially due to the fact that conditional probability is a function in two arguments, and so the number of possibly distinct conditional probabilities is strictly greater than the number of bivalent propositions in any non-trivial finite model.[5] Specifically, a bivalent model based on n worlds generates a Boolean algebra of 2^n bivalent propositions, and so at most 2^n distinct unconditional probabilities. However, the same model can in principle specify a much larger number of conditional probabilities. Excepting the empty set, for each bivalent A consisting of $|A|$ worlds, there are $2^{|A|}$ possibly distinct conditional probability values—one for each subset of A. (The number is smaller than 2^n because, for any two bivalent propositions C and D with $C \cap A = C \cap D$, the conditional probabilities $P_B(C \mid A)$ and $P_B(D \mid A)$ will match.)

5 Hájek's result is a bit stronger: he shows that, in any finite model with more than two worlds the number of actually distinct conditional probabilities will always exceed the number of distinct unconditional probabilities. The result described in the main text shows only that, for any such model, it is possible to define sensible models with one or many unmatched conditional probability values. This result is still quite damning for the conjunction of bivalence with Stalnaker's thesis, and I find that it illuminates the source of the cardinality issue more clearly than Hájek's (1989) original proof. Since every unconditional probability is trivially also a conditional probability, to get from here to Hájek's full result it is necessary to show in addition that every model has at least one unmatched conditional probability.

The model on the left of Figure 2 (cf. 1 above) illustrates using Hájek's (1989) example, with three worlds $\{a,b,c\}$. In this model there are $2^3 = 8$ bivalent propositions, and so at most 8 distinct unconditional probabilities. 7 of these 8 — all but $\varnothing$ – can serve as the second argument to the conditional probability function, and the number of possibly distinct conditional probability values with this conditioning operation depends on their cardinality. Conditioning on $A = \{b,c\}$, for example, generates at most four distinct conditional probabilities, because the numerator of the ratio formula $P_B(C \mid A)$ depends not on C itself but on the intersection of C with A, and there are at most four distinct values of $C \cap A$: (viz., the subalgebra $\{\varnothing, \{b\}, \{c\}, \{b,c\}\}$ generated by A, see Figure 2, right). However, at most 2 of these will be associated with values distinct from 0 and 1, and so conditioning on A generates at most two new conditional probability values.

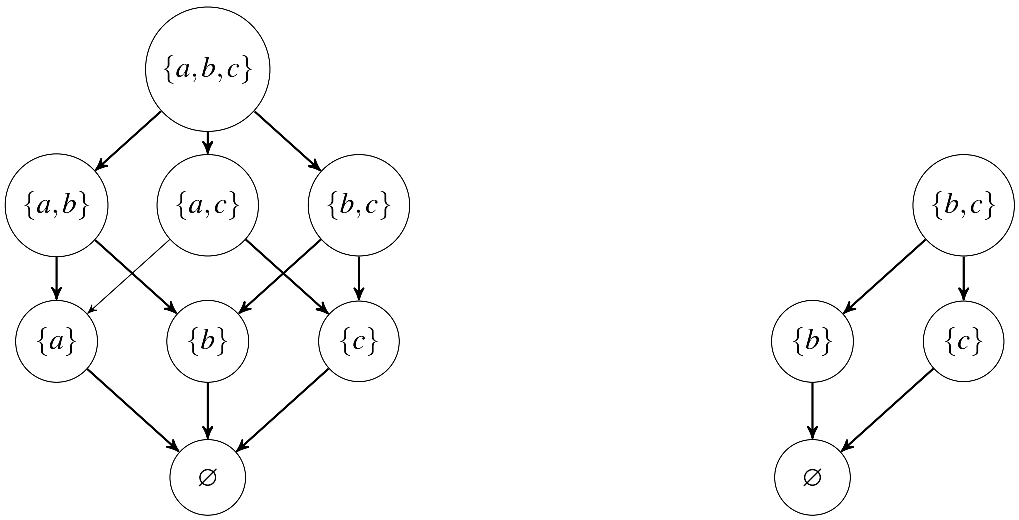

Figure 2 Left: Boolean algebra based on three worlds. Right: Subalgebra associated with conditioning on $\{b,c\}$.

Reasoning in this way, we can calculate that the maximum number of possibly distinct conditional probabilities that can be defined on 3 worlds is

- 1 for the empty set, which gets probability 0 conditioned on anything (except itself);
- 1 for all conditional probabilities of the form $P(A \mid A)$, which are 1 for any $A \neq \varnothing$ unless $P(A) = 0$;

plus

- Conditioning on $\{a,b,c\}$: 2^3 - 2 = 6 (full algebra minus $\varnothing$ and $\{a,b,c\}$ itself)
- Conditioning on $\{a,b\}$: 2^2 - 2 = 2 (subalgebra based on $\{a,b\}$ minus $\varnothing$ and $\{a,b\}$ itself)
- Conditioning on $\{b,c\}$: 2^2 - 2 = 2
- Conditioning on $\{a,c\}$: 2^2 - 2 = 2
- Conditioning on $\{a\}$, $\{b\}$, or $\{c\}$: 2^1 - 2 = 0

There are 14 possibly distinct conditional probability values, but only $2^3 = 8$ bivalent propositions and so at most 8 distinct unconditional probability values—guaranteeing that some models will have wallflowers.

More generally, using the fact that a finite Boolean algebra based on n elements contains $\binom{n}{m}$ distinct sets of cardinality m, each of which generates a subalgebra with 2^m elements, we can calculate that the maximum number of possibly distinct conditional probability values is

$$CP(n) = 2 + \sum_{m=1}^{n} \left[\binom{n}{m} \times (2^m - 2) \right],$$

which is much greater than 2^n except in trivial models. For instance, with $n = 5$ we have $2^n = 32$ but $CP(n) = 182$. With $n = 20$, $CP(n)$ is three orders of magnitude larger than 2^n. These values are the same—$CP(n) = 2^n$—only when n is 1 or 2.

In Hájek's specific example, we have four distinct unconditional probability values ($0, 1/3, 2/3$, and 1) and five distinct conditional probabilities ($0, 1/3, 1/2, 2/3$, and 1). Hájek's example does not exhaust the number of distinct conditional probabilities that *could* be defined on a 3-world model, because the singletons have the same probabilities. In a biased lottery there could well be additional distinct values: for instance, if the tickets have probabilities $.6, .3$, and $.1$ of winning we see all 14 possibly distinct unconditional probabilities.

In contrast, the trivalent semantics encounters no problem with Hájek's Wallflower example. Instead we have (for bivalent A, C and with $P_T(A) > 0$)

$$\begin{aligned}
P_T(A \Rightarrow C) &= \frac{P_T(True(A \Rightarrow C))}{P_T(TV(A \Rightarrow C))} \\
&= \frac{P_T(A \wedge C)}{P_T([A \wedge C] \vee [A \wedge \neg C])} \\
&= \frac{1/3}{1/3 + 1/3} \\
&= 1/2,
\end{aligned}$$

as required.

The trivalent semantics avoids the Wallflower result because—as we saw in the intensional definition of trivalent conditionals in section 2.1—it takes two bivalent propositions to specify a trivalent conditional $A \Rightarrow C$. In fact, the number of possibly distinct trivalent conditionals is strictly greater than the number of possibly distinct conditional probabilities. This is because each conditional $A \Rightarrow C$ is associated with a domain of worlds $TV(A \Rightarrow C)$ in which it is defined, corresponding to the set of worlds in which the antecedent is true. (I am simplifying, inessentially, by assuming that A and C are bivalent.) If this domain has cardinality m, there are 2^m logically distinct conditionals that can be defined with this antecedent, corresponding to all of the elements of the subalgebra generated by A. So, $TC(n)$—the number of logically distinct trivalent conditionals that can be defined in a model with n worlds—is

$$TC(n) = \sum_{m=1}^{n} \left[\binom{n}{m} \times 2^m \right].$$

$TC(n)$ is strictly greater than $CP(n)$ as defined above, for all $n > 1$. This is because, for example, the conditionals *If A then A* and *If B then B* are logically distinct in the trivalent semantics when $A \neq B$ since they have different domains, even though their probabilities are (if defined) both equal to 1. For instance, when

$n = 5$ we have $2^n = 32$, $CP(n) = 182$, $TC(n) = 242$. With $n = 20$, we have $2^n \approx 10^6$, $CP(n) \approx 3.484 \times 10^9$, and $TC(n) \approx 3.486 \times 10^9$. (In general, the number of logically distinct trivalent conditionals definable on n worlds exceeds the maximum number of possibly distinct conditional probabilities by $2 \times (2^n - 2)$.)

Trivalent conditionals provide a better match than bivalent conditionals to the cardinality properties of conditional probabilities: they provide room for a theory that validates Stalnaker's thesis to provide all of the distinct conditional probabilities that it needs to. A bivalent theory, in contrast, forces us to search for ways to jam a huge number of conditional probabilities into a much-too-small space of bivalent propositions.

13 Hájek 2011

Hájek (2011: 11-12) produces a proof showing that no update operation P that satisfies certain obvious desiderata can enforce Stalnaker's thesis except in trivial probability models. Writing P^A for the update of P by A:

- Boldness: If $P(A) > 0$, then $P^A(A) = 1$.
- Moderation: If B entails A and $P(B) > 0$, then $P^A(B) > 0$.

Conditioning has both of these properties, both in bivalent and in trivalent systems. This allows Hájek to show, for the bivalent system, that Stalnaker's thesis is incompatible with update by conditioning along with any other plausible update operation (on the assumption that non-trivial probability models are possible, of course).

Hájek's proof proceeds by cases. In the first part he considers the possibility that $A \wedge C \wedge \neg(A \Rightarrow C)$ is not a contradiction. This case is obviously ruled out by the trivalent semantics (or any semantics that satisfies the highly plausible condition of strong centering). The more interesting case involves the possibility that $A \wedge C \wedge \neg(A \Rightarrow C)$ is a contradiction. Noting that non-triviality implies $P_B((A \Rightarrow C) \wedge \neg(A \wedge C)) > 0$, Hájek points out that update with $X = \neg(A \wedge C)$ is legitimate according to his desiderata, and should yield a revised distribution P^X that continues to assign non-zero probability to $A \Rightarrow C \wedge \neg(A \wedge C)$. $P^X(A \Rightarrow C)$ also has non-zero probability after update with X, since $A \Rightarrow C \wedge \neg(A \wedge C)$ entails $A \Rightarrow C$.

The non-zero probability of the conditional contradicts Stalnaker's thesis, since $P^X(C \mid A) = P^X(A \wedge C)/P^X(A)$ **is** equal to 0 given that $P^X(A \wedge C) = 0$ and the ratio is defined.

This reasoning relies crucially on bivalence: if we assume the trivalent system and the associated definition of conditional probability, we have a conditioning operation that satisfies Hájek's desiderata but for which his proof does not go through. In the trivalent system, $P_T((A \Rightarrow C) \wedge \neg(A \wedge C))$ is either 0 or undefined (the latter when $P_T(\neg A \wedge C) = 0$). Conditioning any non-trivial distribution of this form on $\neg(A \wedge C)$ yields

$$\begin{aligned} P_T(A \Rightarrow C \mid \neg(A \wedge C)) &= \frac{P_T(A \Rightarrow C \wedge \neg(A \wedge C))}{TV(A \Rightarrow C \wedge \neg(A \wedge C)) \wedge \neg(A \wedge C)} \\ &= \frac{0}{P_T(A \wedge \neg C)} \\ &= 0. \end{aligned}$$

(The denominator is $P_T(A \wedge \neg C)$ because $A \Rightarrow C$ has a truth-value only in A-worlds, which together with $\neg(A \wedge C)$ implies $A \wedge \neg C$.) Yet the trivalent system enforces Stalnaker's thesis, and its conditioning operation is bold and moderate:

- Boldness: If $P_T(A) > 0$, then $P_T(A \mid A) = 1$.
- Moderation: If B entails A and $P_T(B) > 0$, then

$$
\begin{aligned}
P_T(B \mid A) &= \frac{P_T(B \wedge A)}{P_T(TV(B \wedge A) \wedge True(A))} \\
&= \frac{P_T(B)}{P_T(A)} \\
&> 0.
\end{aligned}
$$

Given this, it is tempting to construe Hájek's proof as showing that *three* obvious desiderata—Boldness, Moderation, and Stalnaker's thesis—are jointly incompatible with bivalence.

14 Conclusions

The trivalent semantics has not really been taken seriously in much of the literature on conditionals. Nevertheless, it has a number of attractions. It has considerable motivation from intuitions and experimental results around conditional bets, and it allows us to derive conditional restriction of quantifiers and a variety of other operators—including probability operators and quantificational adverbs—without further ado (cf. Belnap 1970; Huitink 2008).

This paper explored the ways in which trivalence allows us to avoid a variety of triviality results that plague bivalent theories, maintaining the empirically well-motivated link between the probabilities of conditionals and conditional probabilities in a straightforward way. The crucial feature of trivalent probability is the fact that the standard ratio definition is not appropriate in trivalent models. From here, it is tempting to view the triviality results as *reductio* arguments against bivalence: since Stalnaker's thesis is correct, they suggest that we should give up the assumption that conditionals must somehow be assigned truth-values when their antecedents are false.

The main alternative is to suppose that conditionals simply lack truth-conditions altogether, as Adams (1975); Edgington (1995) argue. While treating conditionals as semantically special in this way does block certain triviality results, this price deserves Lewis' 'exorbitant' label far more than the trivalent account does. The trivalent theory eliminates the need for special pleading around compounds, nestings, and embeddings of conditionals that goes with non-propositionalism. In addition, when supplemented with a plausible account of assertion for trivalent sentences it may do justice to many of the intuitions around 'supposition' that motivate the non-propositional account (compare McDermott 1996; Rothschild 2014b; Baratgin & Politzer 2016). I conclude that the trivalent semantics and probability is not only an option worth taking seriously, but a clear candidate for the most empirically adequate theory of indicative conditionals available in light of its effortless handling of a variety of deep empirical and technical problems for standard accounts of conditionals.

This is not to say that important challenges do not remain for trivalent theorists. In particular, there is still need for a detailed and linguistically motivated theory of compounds and nestings of conditionals within this framework. In addition, there is important formal and empirical work to be done comparing the trivalent semantics with bivalent (e.g., Khoo to appear) and many-valued (e.g., Stalnaker & Jeffrey 1994; Kaufmann 2009) theories that also avoid at least some triviality results. Finally, the trivalent semantics owes us an account of counterfactuals that respects their morphosyntactic relationship to indicatives and addresses recent triviality results for counterfactuals (Williams 2012; Briggs 2017; Schwarz 2018).

References

Adams, Ernest W. 1975. *The logic of conditionals: An application of probability to deductive logic*. Springer.

Baratgin, Jean, David Over & Guy Politzer. 2014. New psychological paradigm for conditionals and general de finetti tables. *Mind & Language* 29(1). 73–84.

Baratgin, Jean & Guy Politzer. 2016. Logic, probability and inference: a methodology for a new paradigm. *Cognitive unconscious and human rationality* 119–142.

Baratgin, Jean, Guy Politzer, David E Over & Tatsuji Takahashi. 2018. The psychology of uncertainty and three-valued truth tables. *Frontiers in psychology* 9. 1479.

Belnap, Nuel D. 1970. Conditional assertion and restricted quantification. *Noûs* 4(1). 1–12.

Benferhat, Salem, Didier Dubois & Henri Prade. 1997. Nonmonotonic reasoning, conditional objects and possibility theory. *Artificial Intelligence* 92(1-2). 259–276.

Bradley, Richard. 2000. A preservation condition for conditionals. *Analysis* 60(3). 219–222.

Bradley, Richard. 2002. Indicative conditionals. *Erkenntnis* 56(3). 345–378.

Bradley, Richard. 2007. A defence of the Ramsey test. *Mind* 116(461). 1–21.

Briggs, R. A. 2017. Two interpretations of the Ramsey test. In C. Hitchcock H. Beebee & H. Price (eds.), *Making a Difference: Essays on the Philosophy of Causation*, 33–57. Oxford University Press.

Cantwell, John. 2006. The laws of non-bivalent probability. *Logic and Logical Philosophy* 15(2). 163–171.

Cantwell, John. 2008. Indicative conditionals: Factual or epistemic? *Studia Logica* 88(1). 157–194.

Chemla, Emmanuel & Paul Egré. 2019. From many-valued consequence to many-valued connectives. *Synthese* To appear.

Chemla, Emmanuel, Paul Égré & Benjamin Spector. 2017. Characterizing logical consequence in many-valued logic. *Journal of Logic and Computation* 27(7). 2193–2226.

Cooper, William S. 1968. The propositional logic of ordinary discourse. *Inquiry* 11(1-4). 295–320.

Douven, Igor. 2016. On de Finetti on iterated conditionals. In *Computational models of rationality: Essays dedicated to gabriele kern-isberner on the occasion of her 60th birthday*, 265–279. College Publications.

Douven, Igor & Sara Verbrugge. 2013. The probabilities of conditionals revisited. *Cognitive Science* 37(4). 711–730.

Edgington, Dorothy. 1995. On conditionals. *Mind* 104(414). 235–329.

Edgington, Dorothy. 2014. Indicative conditionals. In Edward N. Zalta (ed.), *The stanford encyclopedia of philosophy*, Metaphysics Research Lab, Stanford University winter 2014 edn.

Égré, Paul & Mikael Cozic. 2016. Conditionals. In Maria Aloni & Paul Dekker (eds.), *Cambridge handbook of formal semantics*, Cambridge University Press.

Egré, Paul, Lorenzo Rossi & Jan Sprenger. 2019. De Finettian logics of indicative conditionals. To appear in *Journal of Philosophical Logic*.

Evans, Johathan St. B. T. & David E. Over. 2004. If.

de Finetti, Bruno. 1936. La logique de la probabilité. In *Actes du congrès international de philosophie scientifique*, vol. 4, 1–9. Hermann Editeurs Paris.

de Finetti, Bruno. 1995. The logic of probability. *Philosophical Studies* 77(1). 181–190.

Fitelson, Branden. 2015. The strongest possible lewisian triviality result. *Thought: A Journal of Philosophy* 4(2). 69–74.

Hájek, Alan. 1989. Probabilities of conditionals – revisited. *Journal of Philosophical Logic* 18(4). 423–428.

Hájek, Alan. 2011. Triviality pursuit. *Topoi* 30(1). 3–15.

Huitink, Janneke. 2008. *Modals, Conditionals and Compositionality*: Radboud University Nijmegen dissertation.

Jeffrey, Richard C. 1963. On indeterminate conditionals. *Philosophical Studies* 14(3). 37–43.
Kaufmann, Stefan. 2009. Conditionals right and left: Probabilities for the whole family. *Journal of Philosophical Logic* 38(1). 1–53. doi:10.1007/s10992-008-9088-0.
Khoo, Justin. to appear. *The meaning of* if. Oxford University Press.
Khoo, Justin & Paolo Santorio. 2018. Lecture notes: Probability of conditionals in modal semantics. Lecture notes, 2018 North American Summer School in Logic, Language, and Information. http://www.justinkhoo.org/KS-NASSLLI2018.pdf.
Kolmogorov, Andrey. 1933. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Julius Springer.
Lewis, David. 1975. Adverbs of quantification. In Edward L. Keenan (ed.), *Formal semantics of natural language*, 178–188. Cambridge University Press.
Lewis, David. 1976. Probabilities of conditionals and conditional probabilities. *Philosophical Review* 85(3). 297–315. doi:10.2307/2184045.
McDermott, Michael. 1996. On the truth conditions of certain 'if'-sentences. *The Philosophical Review* 105(1). 1–37.
Milne, Peter. 1997. Bruno de Finetti and the logic of conditional events. *The British Journal for the Philosophy of Science* 48(2). 195–232.
Milne, Peter. 2003. The simplest Lewis-style triviality proof yet? *Analysis* 63(4). 300–303.
Politzer, Guy, David E Over & Jean Baratgin. 2010. Betting on conditionals. *Thinking & Reasoning* 16(3). 172–197.
Rothschild, Daniel. 2014a. Capturing the relationship between conditionals and conditional probability with a trivalent semantics. *Journal of Applied Non-Classical Logics* 24(1-2). 144–152.
Rothschild, Daniel. 2014b. A note on conditionals and restrictors. In John Hawthorne & Lee Walters (eds.), *Conditionals, probability, and paradox: Themes from the philosophy of dorothy edgington*, Oxford University Press.
Schwarz, Wolfgang. 2018. Subjunctive conditional probability. *Journal of Philosophical Logic* 47(1). 47–66.
Stalnaker, Robert. 1976. Letter to van Fraassen. In W. L. Harper & C. A. Hooker (eds.), *Foundations of probability theory, statistical inference, and statistical theories of science*, vol. 1, 302–306. Reidel.
Stalnaker, Robert & Richard Jeffrey. 1994. Conditionals as random variables. In *Probability and conditionals: Belief revision and rational decision*, 31–46. Cambridge University Press.
Stalnaker, Robert C. 1970. Probability and conditionals. *Philosophy of science* 37(1). 64–80.
Williams, J. Robert G. 2012. Counterfactual triviality: A Lewis-impossibility argument for counterfactuals. *Philosophy and Phenomenological Research* 85(3). 648–670.