What we can learn from how trivalent conditionals avoid triviality

Daniel Lassiter
Stanford University

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 analyze precisely how this modification 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. I suggest that the slew of triviality 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 conditionals somehow be assigned a truth-value in worlds in which their antecedents are false.

Keywords: Indicative conditionals, probability, trivalence

Approximate word count: 5,260 (excluding references)

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 ⇒ 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) proved 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 \( 0 < P(B) < 1 \),

\[
P(A \Rightarrow C) = P(A \Rightarrow C \land B) + P(A \Rightarrow C \land \lnot B) = P(A \Rightarrow C \mid B) \times P(B) + P(A \Rightarrow C \mid \lnot B) \times P(\lnot B)
\]

ii. \( P \) is closed under conditionalization, so that—given equation (1)—\[ P(A \Rightarrow C \mid B) = P(C \mid A \land B) \] for any \( B \).
Triviality follows immediately: taking $B = C$, (i) and (ii) give us

\[
P(A \Rightarrow C) = P(A \Rightarrow C|C) \times P(C) + P(A \Rightarrow C|\neg C) \times P(\neg C) \\
= P(C|A \land C) \times P(C) + P(C|A \land \neg C) \times P(\neg C) \\
= 1 \times P(C) + 0 \times P(\neg C) \\
= P(C),
\]

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, beginning with De Finetti (1989; originally published in 1935) and continuing in philosophy through Jeffrey 1963; Belnap 1970; McDermott 1996; Milne 1997; Cantwell 2008; Huitink 2008; Rothschild 2014a and also in 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, several of which fail for exactly the same reason as Lewis’. 1 Here again the way that the trivalent system avoids triviality sheds light on key moves in the original proofs. 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.

2 Trivalent semantics and probability

In De Finetti’s (1989) 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 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.

1 See Khoo & Santorio 2018 for an extremely valuable recent survey of triviality results.
The trivalent semantics may make various predictions about embedded conditionals, depending on the definitions of other connectives and embedding operators. For the purposes of this paper we need only to assume that the negation of # is #, and that the conjunction of # with anything is #.

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 antecedent 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 \( \text{True} \). We allow that some sentences are undefined at some assignments, i.e., neither \( \text{True}(A) \) nor \( \text{True}(\neg A) \) holds. A unary operator \( TV \) (‘has a truth-value’) is defined as \( TV(A) = \text{True}(A) \lor \text{True}(\neg A) \). A bivalent sentence \( A \) is one that has a truth-value at every assignment. Cantwell (2006, 2008) axiomatises a non-bivalent probability measure \( P_T: \mathcal{L} \rightarrow [0,1] \) 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.
  - \( P_T(\neg A) = 1 - P_T(A) \).
  - \( P_T(A \lor B) = P_T(A) + P_T(B) \) if \( A \) and \( B \) are inconsistent.

- For any \( A \):
  - \( P_T(A) = \frac{P_T(\text{True}(A))}{P_T(TV(A))} \), 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 proves that that an 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

---

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 \( \land \) and \( \lor \) so defined—and more generally how they handle undefinedness—which are disputed by McDermott (1996). That said, trivalent theorists do face substantial challenges from embedded and nested conditionals which still need to be addressed in full detail.
We can now state a simple intensional definition of the trivalent probability measure $P$ which takes as its argument trivalent propositions (i.e., pairs of classical propositions; for readability I will write $\land$ the composition of $J$ everywhere, the third axiom places on bivalent propositions only the trivial requirement that $P_J(W) = 1$; $\land$ $P_J(W, D \cup E) = P_J(W, D) + P_J(W, E)$ if $D \cap E = \emptyset$; $\land$ $P_J(D, E) = \frac{P_J(W, D \cap E)}{P_J(W, D)}$. 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 $(W, D)$ is defined everywhere, the third axiom places on bivalent propositions only the trivial requirement that $P_J(W, E) = P_J(W, W \cap E)/P_J(W, W) = P_J(W, E)/1$. It is a straightforward exercise to show that in the intensional system the composition of $[\cdot]$ with $P_J$ enforces Cantwell’s axioms.
2.2 Conditional probabilities in the trivalent setting

This section draws out the modification of the ratio definition conditional 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 $A$—

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

—but this similarity hides a difference from standard probability that will be important in what follows. As noted above, we assume that a conjunction is undefined if either of its conjuncts is: $1 \land \# = \# \land 1 = \# \land \# = \#$. Therefore, if one or both conjuncts of $C \land A$ is not bivalent, the calculation of the probability of a conjunction depends on the probability that it has a truth-value. From REST-E we have

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

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

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

provided that $P_T(TV(C \land A) \land True(A)) > 0$. The normalizing 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 $D, E, A$)

$$P_T(D \Rightarrow E | A) = \frac{P_T(True(D \Rightarrow E \land A))}{P_T(TV(D \Rightarrow E \land A) \land True(A))}$$

(2)

$$= \frac{P_T(D \land E \land A)}{P_T(D \land A)}.$$  

(3)

The conditional probability of a trivalent conditional thus depends on normalization by the probability of the conjunction of the conditioning proposition and the conditional’s antecedent.

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

$$P_B(D \land B) = P_B(D | B) \times P_B(B)$$

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

$$P_B(D | B) \times P_B(B) = \frac{P_B(D \land B)}{P_B(B)} \times P_B(B)$$

$$= P_B(D \land B)$$
This equivalence does not hold for $P_T$ when $D$ happens to be trivalent. Instead we have (for bivalent $A,C,B$):

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

which does not reduce to $P_T(A \Rightarrow C \land B)$ because of the nonstandard normalization. This difference from standard probability will play a starring role when $D$ is a conditional, which is the crucial case in a number of the triviality proofs that we will examine. In bivalent probability, we would have instead the equivalence

$$(4) \quad P_B(A \Rightarrow C \land B) = P_B(A \Rightarrow C \mid B) \times P_B(B).$$

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.

### 2.3 Stalnaker’s thesis

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

$$P_T(A \Rightarrow C) = \frac{P_T(\text{True}(A \Rightarrow C))}{P_T(\text{TV}(A \Rightarrow C))} = \frac{P_T(A \land C)}{P_T(A)} = P_T(C \mid A).$$

In the trivalent semantics, 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 the axiomatization of probability.

### 3 Warm-up to triviality proofs: Hájek’s (1989) Wallflower result

Consider the model that Hájek (1989) uses to illustrate his ‘Wallflower’ result: a finite probability distribution on $W = \{w_1, w_2, w_3\}$, 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 $w_1$ and $w_2$, and let $C$ be true only at $w_1$. 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.

Hájek’s result depends crucially on bivalence. In the trivalent semantics, we have (for bivalent $A,C$ and with
\[ P_T(TV(A \Rightarrow C)) > 0 \]

\[
P_T(A \Rightarrow C) = \frac{P_T(True(A \Rightarrow C))}{P_T(TV(A \Rightarrow C))} \]
\[
= \frac{P_T(A \land C)}{P_T([A \land C] \lor [A \land \neg C])} \]
\[
= \frac{1}{3} \cdot \frac{2}{3} \]
\[
= \frac{1}{2},
\]

as required.

In effect, the trivalent semantics avoids the Wallflower result because the the set of trivalent propositions in a model with a fixed set of worlds is larger than the set of bivalent propositions. Since conditional probability is a function in two arguments, a model with \( n \) worlds defines \( 2^n \) pairs of bivalent propositions for which one could in principle take as arguments to conditional probability (though some are filtered out, e.g., one cannot condition on a contradiction, and a large number will be trivially mapped to 1 or 0). However, an unconditional measure specifies only \( 2^n \) unconditional probabilities—a much smaller number even when limitations imposed by the probability axioms and the ratio formula are taken into account.

In contrast, it takes two bivalent propositions to specify a trivalent conditional \( A \Rightarrow C \) (as we saw in the intensional definition of trivalent probability in section 2.1). The first specifies the region of logical space where the conditional is defined; the second specifies the subset of that region where the conditional is true. This means that the number of logically distinct trivalent conditionals definable on \( n \) worlds is \( 2^{2n} \), which is on the same order of magnitude as the largest number of possibly distinct conditional probabilities (indeed, strictly larger). There is no pressure to squeeze all of the possible conditional probabilities into a much-too-small space of unconditional probabilities.

## 4 Lewis’s (1976) proof

We know that already that the trivalent semantics avoids Lewis’s (1976) triviality results—but how? The pivotal moment in the proof is the transition from line (5) to (6).

\[
(5) \quad P_B(A \Rightarrow C) = P_B(A \Rightarrow C \land C) + P_B(A \Rightarrow C \land \neg C)
\]
\[
(6) \quad = P_B(A \Rightarrow C | C) \times P_B(C) + P_B(A \Rightarrow C | \neg C) \times P_B(\neg C)
\]
\[
(7) \quad = P_B(C | A \land C) \times P_B(C) + P_B(C | A \land \neg C) \times P_B(\neg C)
\]
\[
(8) \quad = 1 \times P_B(C) + 0 \times P_B(\neg C)
\]

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: \( P_T(A \Rightarrow C \land C) \) does not in general equal \( P_T(A \Rightarrow C | C) \times P_T(C) \). Instead we have (for bivalent \( A, C \), and assuming \( P_T(A \land

\[
7
\]
\( \), \( P_T(A \land \neg C) > 0 \):

\[
P_T(A \Rightarrow C) = P_T(A \Rightarrow C \land C) + P_T(A \Rightarrow C \land \neg C)
\]

\[
= \frac{P_T(A \lor C \land C)}{P_T(A)} + \frac{P_T(A \lor C \land \neg C)}{P_T(A)}
\]

\[
= \frac{P_T(A \lor C)}{P_T(A)} + 0
\]

which does not have any interesting relation to the trivializing term \( P_T(A \Rightarrow C | C) \times P_T(C) \) as it would in bivalent semantics. Indeed, the proof is simply a roundabout way to derive Stalnaker’s thesis. In sum, nothing of interest follows in the trivalent setting from the application of the law of total probability to conditionals.\(^3\)

Like the Wallflower result, Lewis’ proof seems to show that probability and conditionals don’t get along well. I suggest that this is because we have been trying to apply to conditionals rules that are valid only for the special case of bivalent propositions. This diagnosis also covers most of the triviality proofs to be considered in the rest of the paper.

5 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 \( P_B(A | C) = P_B(A \supset C) \) only if \( P_B(A) = 1 \), \( P_B(C | A) = 1 \), or \( P_B(A \land C) = 0 \).

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

\[
P_B(A \Rightarrow C) = P_B(A \Rightarrow C \land A \supset C) + P_B(A \Rightarrow C \land \neg A \supset C)
\]

\[
\geq P_B(A \Rightarrow C \land A \supset C)
\]

\[
= P_B(A \Rightarrow C | A \supset C) \times P_B(A \supset C)
\]

\[
= P_B(A \supset C)
\]

assuming \( P_B(A \land C) > 0 \).

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

\[
P_T(A \Rightarrow C \land A \supset C) = \frac{P_T(A \land C \land A \supset C)}{P_T(A)}
\]

\[
= \frac{P_T(A \land C)}{P_T(A)}
\]

\[
= P_T(A \Rightarrow C)
\]

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 inquiry. It does hold, for example, for a notion of consequence \( \models_T \) where \( A \models_T B \) just in case there is no assignment in which \( A \) is not false and \( B \) is false.)

\(^3\) 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.
6 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 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.

\begin{align*}
(13) & \quad P_B(A \Rightarrow C) = P_B(A \Rightarrow C \land C) + P_B(A \Rightarrow C \land \neg C) \\
(14) & \quad = P_B(A \Rightarrow C \mid C) \times P_B(C) + P_B(A \Rightarrow C \mid \neg C) \times P_B(\neg C) \\
(15) & \quad = 1 \times P_B(C) + 0 \times P_B(\neg C)
\end{align*}

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 $P(A) > 0$.

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

7 Fitelson (2015)

The principle of Import-Export is valid in trivalent semantics:

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

Both sides of the equivalence are true if $A \land B \land C$, false if $A \land B \land \neg C$, 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 $Y$,

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

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 2.2 above).

$$P_T(A \Rightarrow C \mid Y) = \frac{P_T(True(A \Rightarrow C \land Y))}{P_T(True(A \Rightarrow C \land Y))}$$
For bivalent $A, C, Y$, this is equivalent to

$$PT(A \Rightarrow C \mid Y) = \frac{PT(A \land C \land Y)}{PT(A \land Y)}$$

$$= PT(C \mid A \land Y)$$

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 trivalent semantics, and the third is Stalnaker’s Thesis itself, which we have seen is unproblematic in trivalent semantics.

- The $\neg C$-instance: $PT(\neg A \Rightarrow C \mid \neg C) = PT(C \mid A \land \neg C)$, and so (by the trivalent ratio definition)
  $$\frac{PT(\neg A \land C \land \neg C)}{PT(\neg A \land \neg C)} = PT(C \mid A \land \neg C) = 0,$$
  which implies that $PT(\neg A \Rightarrow C \mid \neg C) = 0$.

- The $A \supset C$-instance: $PT(\neg A \Rightarrow C \mid A \supset C) = PT(C \mid A \land A \supset C)$, and so (by the bivalent ratio definition)
  $$\frac{PT(\neg A \land C \land A \supset C)}{PT(\neg A \land A \supset C)} = PT(C \mid A \land A \supset C) = 1.$$

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

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: $PT(\neg A \Rightarrow C \mid \neg C) = PT(C \mid A \land \neg C)$, and so (by the trivalent ratio definition)
  $$\frac{PT(A \land C \land \neg C)}{PT(A \land \neg C)} = PT(C \mid A \land \neg C),$$
  which is trivially true.

- The $A \supset C$-instance: $PT(A \Rightarrow C \mid A \supset C) = PT(C \mid A \land A \supset C)$, and so (by the trivalent ratio definition)
  $$\frac{PT(A \land C \land A \supset C)}{PT(A \land A \supset C)} = PT(C \mid A \land A \supset C),$$
  which is also trivially true.

- The tautology: if we substitute a tautology for $Y$ we get Stalnaker’s Thesis: $PT(\neg A \Rightarrow C) = PT(C \mid A)$, 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.
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 $A, B, C$: $P_B(A \Rightarrow C) = P_B(C | A)$.
- 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 $P_B(A \Rightarrow C / \neg C) = P_B(C / \neg A)$. This consequence is essentially due to bivalence: conditionals obviously do not entail the negation of the antecedents, and so (in the bivalent setting) the probability of the conditional must somehow be distributed among the worlds where the antecedent is false. Using the law of total probability, the bivalent ratio formula, and strong centering we get

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

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

$$= 1 \times P_B(A) + P_B(A \Rightarrow C \wedge \neg A).$$

This is, I suggest, where the trouble lies. Substituting $P_B(C | A)$ for $P_B(A \Rightarrow C)$ per Stalnaker’s thesis we now have

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

and so

$$P_B(A \Rightarrow C \wedge \neg A) = P_B(C | A) - P_B(A)$$

As a result, bivalence together with the 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 the trivial case that $P_B(C | A) = P_B(A)$, which holds only if $P_B(C | A) = 1$. But this implies, via Stalnaker’s thesis, that the conditional itself has probability 0 or 1. In a bivalent system that satisfies Stalnaker’s assumptions, any non-extreme assignment of probability to a conditional will have to assign some probability to the conjunction of $A \Rightarrow C$ with the negation of its antecedent.

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 the conditional $A \Rightarrow C$ cannot have nonzero probability conditional on $\neg A$, because it is not defined at $\neg A$-worlds.

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

which is either 0 or undefined.
Hájek (2011) 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 \( A \land C \land \neg(A \Rightarrow C) \) 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 \( A \land C \land \neg(A \Rightarrow C) \) is a contradiction, as in the trivalent semantics. Noting that non-triviality implies \( P_T((A \Rightarrow C) \land \neg(A \land C)) > 0 \) (as we showed in the last section), Hájek points out that update with \( X = \neg(A \land C) \) is legitimate according to his desiderata, and should yield a revised distribution \( P^X \) that continues to assign non-zero probability to \( A \Rightarrow C \land \neg(A \land C) \).

The non-zero probability of the conditional contradicts Stalnaker’s thesis, since \( P^X(C \mid A) = P^X(A \land C)/(P^X(A)) \) is equal to 0 given that \( P^X(A \land C) = 0 \) 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 semantics, \( P_T((A \Rightarrow C) \land \neg(A \land C)) \) is either 0 or undefined (the latter when \( P_T(\neg A \land C) = 0 \)). Conditioning any non-trivial distribution of this form on \( \neg(A \land C) \) yields

\[
P_T(A \Rightarrow C \mid \neg(A \land C)) = \frac{P_T(A \Rightarrow C \land \neg(A \land C))}{TV(A \Rightarrow C \land \neg(A \land C)) \land \neg(A \land C)} = \frac{P_T(A \land \neg C)}{0} = 0.
\]

(The denominator is \( P_T(A \land \neg C) \) because \( A \Rightarrow C \) has a truth-value only in \( A \)-worlds, which together with \( \neg(A \land C) \) implies \( A \land \neg C \).) 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

\[
P_T(B \mid A) = \frac{P_T(B \land A)}{P_T(TV(B \land A) \land True(A))} = \frac{P_T(B)}{P_T(A)} > 0.
\]
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.

10 Bradley (2000)

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

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

The proof again depends on reasoning about the probability of the conjunction \( A \Rightarrow C \land \neg A \), which is non-trivial in bivalent systems but trivially 0 or undefined in the trivalent system.

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

\[
P_T(A \Rightarrow C) = \frac{P_T(A \land C)}{P_T(A)} = \frac{0}{P_T(A)},
\]

with the result that Bradley’s Preservation Condition is a theorem.

11 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 \textit{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). 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 (G. Williams 2012; Briggs 2017; Schwarz 2018).

References


