Triviality Results, Conditional Probability, and Restrictor Conditionals

Jonathan Vandenburgh
Northwestern University
jonathanvandenburgh2021@u.northwestern.edu

October 19, 2020

Abstract

Conditional probability is often used to represent the probability of the conditional. However, triviality results suggest that the thesis that the probability of the conditional always equals conditional probability leads to untenable conclusions. In this paper, I offer an interpretation of this thesis in a possible worlds framework, arguing that the triviality results make assumptions at odds with the use of conditional probability. I argue that these assumptions come from a theory called the operator theory and that the rival restrictor theory can avoid these problematic assumptions. In doing so, I argue that recent extensions of the triviality arguments to restrictor conditionals fail, making assumptions which are only justified on the operator theory.

The most natural way to calculate the probability of a conditional sentence 'If A, then B' is to use the conditional probability of B given A. This is defined using the ratio formula: the probability of B given A, denoted $\Pr(B|A)$, is the probability that both A and B are true, normalized by the probability that A is true:

$$\Pr(B|A) = \frac{\Pr(A \wedge B)}{\Pr(A)}.[1]$$

Conditional probability provides intuitive probability assignments in standard examples in probability and statistics textbooks: the probability that, if the die I rolled lands on an odd number, then it lands on 3 is 1/3; the probability that, if John is an adult male, he is taller than 176 cm is 1/2; the probability that a second coin will land heads, given that the first coin also landed heads, is 1/2. Extensive psychological evidence further supports the claim that people compute the probabilities of conditionals using conditional probability.[2]

[1] This definition only applies if $\Pr(A) \neq 0$. Throughout this paper, I will assume that the antecedents of conditionals have non-zero probability.

[2] For experimental evidence, see Hadjichristidis et al. (2001), Oberauer and Wilhelm (2003), Over and Evans (2003), Evans et al. (2007), Oberauer et al. (2007), Over et al. (2007), and Douven and Verbrugge (2010). This evidence supports the conclusion that conditional probability provides the best predictor of how people evaluate the probability of a

This motivates what is often called 'The Thesis' in the literature: if we abbreviate the conditional sentence 'If A, then B' as $A \to B$, The Thesis states that the probability of a conditional, $\Pr(A \to B)$, is equal to the conditional probability $\Pr(B|A)$:

$$\Pr(A \to B) = \Pr(B|A).$$

This thesis most famously received support in philosophy from Adams (1975) and Stalnaker (1970). However, it is widely acknowledged that The Thesis cannot be true, or at least is inconsistent with otherwise appealing principles of conditional semantics and probability theory, due to the triviality results. The first and most famous triviality result appeared in Lewis (1976), where Lewis argued, making mild assumptions about the logic of conditionals, that a probability assignment satisfying The Thesis must be trivial in the sense that $\Pr(A \to B) = \Pr(B)$. Lewis's triviality result follows from three assumptions:

1. the law of total probability $\Pr(B) = \Pr(B|A)\Pr(A) + \Pr(B|\neg A)\Pr(\neg A)$,
2. the assignment $\Pr(A \to B|B) = 1$, and
3. the assignment $\Pr(A \to B|\neg B) = 0$.

Lewis uses the law of total probability to expand $\Pr(A \to B)$ into

$$\Pr(A \to B) = \Pr(A \to B|B)\Pr(B) + \Pr(A \to B|\neg B)\Pr(\neg B)$$

and then uses the two conditional assignments to compute that this equals

$$1 * \Pr(B) + 0 * \Pr(\neg B) = \Pr(B).$$

This produces the triviality result, that $\Pr(A \to B) = \Pr(B)$, showing that the probability assignment to conditionals is trivial. This conclusion is clearly problematic for the thesis that $\Pr(A \to B) = \Pr(B|A)$. If we consider the example of rolling a fair die and let X be the proposition 'the die lands on an odd number' and Y be the proposition 'the die lands on 3', then $\Pr(Y|X) = 1/3$ but $\Pr(Y) = 1/6$. For someone committed to The Thesis, an explanation is in order for why Lewis's argument cannot be correct.

Many authors seeking to hold onto The Thesis have questioned the semantic assumptions behind the triviality arguments. Adams (1975) and Edgington (1995) argue that conditionals do not have truth values, Bradley (2002) argues that conditionals sometimes have undefined truth values, and Kaufmann (2001) and Rothschild (2014) argue that conditionals sometimes take

conditional. However, the use of conditional probability is not universal: Over and Evans (2003), for example, found that approximately 50% of people used conditional probability, but over 40% used an alternative heuristic, the probability of the conjunction of the antecedent and consequent $\Pr(A \wedge B)$. Nonetheless, conditional probability repeatedly appears as the most commonly used heuristic, and Evans et al. (2007) provides evidence that those using conditional probability performed better on a general intelligence test than those using conjunctive probability.

on an ‘intermediate’ truth value. Other authors argue, contrary to assumptions from the triviality arguments, that the conditional is context-sensitive (Van Fraassen (1976), Bacon (2015), Khoo (2016)). Finally, Egré and Cozic (2011) argue against the assumption that the conditional is a ‘binary, proposition-forming connective (p. 22),’ arguing instead for the adoption of Kratzer’s (1986, 2012) restrictor theory of the conditional.

In this paper, I argue that the triviality arguments rely on a problematic assumption of conditional semantics called the operator theory. I begin by offering an interpretation of The Thesis in a probabilistic formalism grounded in possible worlds (§1). I argue that the triviality results of Lewis (§2) and Bradley (§3) both make assumptions which are at odds with this formulation of The Thesis. In §4, I argue that these problematic assumptions follow from an operator theory of the conditional. I argue that the operator theory directly leads to triviality, following Hájek’s (1989, 2012) ‘Wallflower Argument’ and Egré and Cozic’s undefinability argument, and that the restrictor theory offers an alternative to the operator theory. In §5, I argue that recent arguments to extend triviality results to the restrictor theory, such as that of Charlow (2016), borrow assumptions from the operator theory which an advocate of the restrictor theory need not accept. Many of the conclusions of this paper agree with the restrictor theory of Egré and Cozic (2011), but this paper goes beyond their account by (1) presenting the responses to triviality as a consequence of The Thesis itself rather than the restrictor theory explicitly, (2) highlighting a problematic assumption in Bradley’s triviality result, and (3) responding to recent work aimed directly at establishing triviality for restrictor conditionals.

1 Formulating The Thesis

In possible world semantics, propositions are represented by the set of worlds (or possible situations) in which the proposition is true.[3] Probabilistic language builds on the assumption that different possible situations can be assigned likelihoods of occurring, and propositions are more or less likely to be true based on the relative likelihoods of these situations. We can formalize this intuition by assuming that the set of possible worlds Ω has a probability distribution Pr defined over worlds which assigns a likelihood to the different possible worlds. Formally, a probability distribution over worlds is a tuple $(\Omega, \Sigma, \mathrm{Pr})$ where Ω is the relevant set of possible worlds, Σ is some sigma-algebra on Ω representing which sets of possible worlds are measurable, and Pr is a probability measure on Σ; I will call such a tuple a *probabilistic context*.

When Ω is finite, as will be the case in all examples in this paper, we can ignore Σ and think of Pr as assigning a probability $\mathrm{Pr}(\omega)$ to each world ω in Ω. It is sufficient for a finite (Ω, Pr) to be a probabilistic context that:

(i) for all $\omega \in \Omega$, $\mathrm{Pr}(\omega) \in [0, 1]$

(ii) $\sum_{\omega \in \Omega} \mathrm{Pr}(\omega) = 1$.

[3]For more on possible world (intensional) semantics, see Von Fintel and Heim (2011).

A probabilistic context allows us to calculate the probability that a proposition A is true. Let $\mathcal{A}$ be the set of propositions: for any proposition A, A is either true or false in each world, so it determines an indicator function $\chi_A : \Omega \to \{0, 1\}$ which assigns $\chi_A(\omega) = 1$ if A is true in ω and $\chi_A(\omega) = 0$ if A is false in ω for each $\omega \in \Omega$. We can then calculate the probability that A is true in a probabilistic context (Ω, Pr):

$$\mathrm{Pr}(A) = \sum_{\omega \in \Omega} \mathrm{Pr}(\omega)\chi_A(\omega).$$

This is just the expected value of the random variable χ_A over the space of possible worlds: it yields the expected probability that A is true, given that ω is chosen from Ω according to Pr.[4]

We can see how this framework works in a specific example: tossing a fair six-sided die. The probabilistic context used here involves six possible states of the world, $\{\omega_1, ..., \omega_6\}$, corresponding to rolling the numbers $1, ..., 6$, each with equal probability $1/6$ of occurring. We can then evaluate the probability of any proposition involving these outcomes using the above formula. For example, if X is the proposition that the die lands on an odd number, the probability of X is given by: $\mathrm{Pr}(X) = \sum_{\omega \in \Omega} \mathrm{Pr}(\omega)\chi_X(\omega) = \mathrm{Pr}(\omega_1) + \mathrm{Pr}(\omega_3) + \mathrm{Pr}(\omega_5) = 1/6 + 1/6 + 1/6 = 1/2$. This gives the intuitive result that the probability of a die landing on an odd number is $1/2$.

The best cases for applying this framework are those where a commonly accepted statistical model determines the set of outcomes and the probability distribution: this includes basic examples from statistics like games of chance, quantitative datasets in fields like medicine, political science and economics (i.e., human height and weight statistics), and theoretical models for expected distributions, like particle decay or asset prices. However, there is good reason to believe this probabilistic model extends beyond these cases: the psychological evidence cited above supports the view that people ascribe coherent probabilities to different outcomes, even when real-world knowledge is involved, and this kind of probabilistic analysis underlies the Bayesian approach to epistemology.[5]

So far, we have only considered assigning probabilities to non-conditional propositions, and we face a choice in determining how to model probabilities of conditionals. One natural thought is to treat the conditional as no different from other propositions, computing its probability as above, since we have

[4] When Ω is infinite, an assignment $\mathrm{Pr} : \Omega \to [0, 1]$ satisfying (i) and (ii) is not sufficient to generate a probabilistic context $(\Omega, \Sigma, \mathrm{Pr})$. Here, we need Σ to be a sigma-algebra of subsets of Ω and $\mathrm{Pr} : \Sigma \to [0, 1]$ to satisfy (i) $\mathrm{Pr}(\Omega) = 1$ and (ii) $\forall A_i \in \Sigma, \mathrm{Pr}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathrm{Pr}(A_i)$. In the infinite case, we can calculate $\mathrm{Pr}(A) = \int_{\Omega} \chi_A(\omega) \, \mathrm{Pr}(\omega)$.

[5] Bayesian epistemology assumes that there is a field of propositions $\mathcal{A}$ and that people have a credence function $P : \mathcal{A} \to [0, 1]$ such that (i) for every tautology A, $P(A) = 1$ and (ii) for mutually exclusive A and B, $P(A \vee B) = P(A) + P(B)$. It is not hard to see that, if every proposition in $\mathcal{A}$ is a set of possible worlds, a probabilistic context $(\Omega, \Sigma, \mathrm{Pr})$ induces a credence assignment $\mathrm{Pr} : \mathcal{A} \to [0, 1]$ which satisfies the axioms of Bayesian epistemology. For more on Bayesian epistemology, see Hartmann and Sprenger (2010).

a number of semantic theories providing possible world truth conditions for the conditional. One problem with this approach is that it will never be able to explain the evidence from the introduction that people compute the probability of conditionals as conditional probability; in §4, I will refer to this as the operator theory and show how it leads to the triviality results. Another option, and the one I will follow here, involves taking The Thesis as primitive. On this view, the probability of a conditional $A \to B$, $\Pr(A \to B)$, is directly associated with the conditional probability $\Pr(B|A)$, so $\Pr(A \to B) = \Pr(B|A) = \frac{\Pr(A \wedge B)}{\Pr(A)}$.[6]

Note that the conditional probability $\Pr(-|A)$, in attributing a probability to all propositions B over Ω, defines a new probability distribution over Ω. This means that $\Pr(-|A)$ associates a probability to all measurable subsets in Σ and obeys the Kolmogorov axioms for probability distributions.[7] We can see this clearly in the case of a finite Ω: for any $\omega \in \Omega$, we get $\Pr(\omega|A) = \frac{\Pr(\omega \wedge A)}{\Pr(A)} = \frac{\Pr(\omega)\chi_A(\omega)}{\Pr(A)}$, and we can verify that this satisfies the two conditions introduced in the previous section. To simplify notation, we will use $\Pr_A$ to refer to the conditional probability distribution $\Pr(-|A)$.

In the formulation of The Thesis, we have assumed that A and B are non-conditional propositions, representing measurable subsets of Ω. There are many challenges to extending The Thesis to include more complex propositions incorporated in a conditional sentence, such as modals, probability operators, or other conditionals. Here we discuss one such case: right-nested conditionals. For a nested conditional $A \to (B \to C)$, it is sometimes thought that repeated iterations of The Thesis by itself cannot offer a prediction for the interpretation of the probability $\Pr(A \to (B \to C)$. Applying The Thesis to the outer conditional tells us $\Pr(A \to (B \to C) = \Pr(B \to C|A)$, but then it appears that there is no easy way to reduce this expression further. This is because applying The Thesis twice would yield an expression like $\Pr((C|B)|A)$, which is not mathematically well-defined. However, thinking of conditional probability in terms of the new distribution $\Pr_A$ on Ω can resolve this issue. The Thesis states that $\Pr(A \to B) = \Pr_A(B)$, where $\Pr_A$ is a new distribution over Ω. When the consequent is also a conditional, we get that $\Pr(A \to (B \to C)) = \Pr_A(B \to C) = \Pr_A(C|B)$. This is well defined: $\Pr_A$ is a distribution over Ω and B is a proposition defined over Ω, so we can define the conditional probability distribution $\Pr_A(-|B)$.

Note that this leads to the principle known as probabilistic export-import (PEI): according to The Thesis, $\Pr(A \to (B \to C)) = \Pr_A(B \to C) =$

[6] Note that this proposal leaves unspecified how we interpret the expression $\Pr(A \to B)$. Two common approaches to this are found in Adams (1975), where the probability of a conditional represents the degree to which a conditional is assertible, and Stalnaker (1970), where the probability of a conditional is the probability that the conditional proposition is true; this question will be addressed in greater depth in §4.

[7] For a more detailed discussion, see Capinski and Kopp (2013).

$\Pr_A(C|B)$, and

$$\Pr_A(C|B) = \frac{\Pr_A(B \wedge C)}{\Pr_A(B)} = \frac{\Pr(A \wedge B \wedge C)}{\Pr(A)} * \frac{\Pr(A)}{\Pr(A \wedge B)}$$

$$= \frac{\Pr(A \wedge B \wedge C)}{\Pr(A \wedge B)} = \Pr((A \wedge B) \to C).$$

This shows that, if we permit multiple applications of The Thesis to right-nested conditionals, PEI holds: $\Pr(A \to (B \to C)) = \Pr((A \wedge B) \to C)$. It is worth defending this principle in greater depth because it has been controversial in the literature on triviality: Kaufmann (2001), Douven and Verbrugge (2013), and Fitelson (2014) all argue that the triviality arguments rely on PEI being true; Kaufmann and Douven and Verbrugge recommend rejecting PEI in response to this. These arguments against PEI derive triviality from the combination of PEI, The Thesis and the operator theory (discussed below), concluding that we should reject PEI. However, this paper and Adams (1975) show that one can maintain PEI while avoiding triviality by denying the operator theory, as recommended both here and by Adams. Regarding the status of PEI, there have been some attempts to pose counterexamples, though these counterexamples are contested.[8] Douven and Verbrugge offer empirical evidence against PEI, providing experimental results that suggest that people ascribe different probabilities to the propositions 'If A, then if B, C' and 'If A and B, then C.' van Wijnbergen-Huitink et al. (2015), however, provide evidence in favor of PEI, suggesting that people report similar probabilities and truth conditions for right-nested conditionals in both iterated and imported form.[9] Since PEI is a plausible extension of the logical export-import principle with experimental support and no clear grounds for rejection, we shouldn't be concerned that it is a consequence of the Bayesian approach.

2 Lewis's Triviality Result

The last section introduced The Thesis, $\Pr(A \to B) = \Pr_A(B)$, for simple conditionals and right-nested conditionals. For right-nested conditionals, we

[8]Kaufmann attempts to provide a counterexample to PEI: consider 'If a match lights, then if you strike it, it will light' and 'If a match lights and you strike it, then it will light.' Kaufmann argues that the latter is necessarily true, but the former may not be true: if you throw a match into a fire, then it is not necessarily true that 'If you strike it, it will light.' Khoo and Mandelkern (2018) offer a compelling response to this, arguing that Kaufmann ignores the temporal properties introduced by 'will': this leads Mandelkern (2018, p. 4) to conclude 'the subsequent literature has not produced a convincing counterexample to [Export-Import], at least for indicative conditionals.'

[9]van Wijnbergen-Huitnik et al note that the conditionals in their experiments rely on less real-world knowledge than those used in Douven and Verbrugge's experiment, offering a potential explanation for the divergence in results. Given the complexity of evaluating nested conditionals, one would expect people to assign conditional probabilities more accurately in simpler scenarios involving less real-world knowledge.

saw that allowing multiple iterations of The Thesis entails probabilistic export-import (PEI). This puts us in a position to evaluate Lewis's triviality result.

Recall that Lewis's triviality result proceeds as follows: Lewis uses the law of total probability to expand $\Pr(A \to B)$ into

$$\Pr(A \to B) = \Pr(A \to B|B)\Pr(B) + \Pr(A \to B|\neg B)\Pr(\neg B)$$

and then computes that this equals

$$1 * \Pr(B) + 0 * \Pr(\neg B) = \Pr(B).$$

This produces the triviality result, that $\Pr(A \to B) = \Pr(B)$. Recall also that the conclusion of this argument does not hold. Let X be the event of rolling an odd number on a fair die and Y be the event of rolling a three. Evaluating conditional probability, we see that $\Pr(X \to Y) = \frac{\Pr(X \wedge Y)}{\Pr(X)} = \frac{1/6}{1/2} = 1/3$, but $\Pr(Y) = 1/6$. Since $\Pr(X \to Y) \neq \Pr(Y)$, the conclusion of Lewis's argument clearly does not hold.

If the conclusion of triviality does not apply to real assignments of conditional probability, that means something must have gone wrong in Lewis's argument. To figure out where the problem arises, we can reconsider his three assumptions: (1) the law of total probability, (2) the assignment $\Pr(A \to B|B) = 1$, and (3) the assignment $\Pr(A \to B|\neg B) = 0$. Note that the ability to iterate The Thesis for right-nested conditionals entails that (2) and (3) are true: $\Pr_B(A \to B) = \Pr(B|A \wedge B) = 1$ and $\Pr_{\neg B}(A \to B) = \Pr(B|A \wedge \neg B) = 0$. This means that the error must lie with Lewis's application of the law of total probability. In fact, I will argue that Lewis's conditional version of the law of total probability does not follow from the axioms of probability theory and is instead a substantial assumption beyond what would be plausible to an advocate of The Thesis.

Before discussing how the law of total probability interacts with conditional probability, it will be useful to revisit the rule for non-conditional propositions. For any A and B, the law of total probability states that $\Pr(B) = \Pr(B|A)\Pr(A) + \Pr(B|\neg A)\Pr(\neg A)$. We can prove this using the ratio definition of conditional probability:

$$\Pr(B|A)\Pr(A) + \Pr(B|\neg A)\Pr(\neg A) = \frac{\Pr(B \wedge A)}{\Pr(A)}\Pr(A) + \frac{\Pr(B \wedge \neg A)}{\Pr(\neg A)}\Pr(\neg A)$$

$$= \Pr(B \wedge A) + \Pr(B \wedge \neg A) = \Pr(B).$$

Lewis's argument, however, relies on a conditional law of total probability: $\Pr(A \to B|C)\Pr(C) + \Pr(A \to B|\neg C)\Pr(\neg C)$. This rule, however, is not a simple consequence of the axioms of probability theory as the non-conditional rule is. Instead, Lewis's conditional law of total probability relies on the extra assumption of the operator theory, which I discuss in §4 and argue is incompatible with The Thesis.

We can use The Thesis to derive an alternative expression for the law of total probability for conditionals.[10] Since B is a non-conditional proposition, the law of total probability must hold with respect to the conditional distribution $\Pr_A$, so for any C:

$$\Pr_A(B) = \Pr_A(B|C)\Pr_A(C) + \Pr_A(B|\neg C)\Pr_A(\neg C).$$

We can then apply The Thesis for non-conditional propositions to re-write this:

$$\Pr_A(B) = \Pr_A(C \to B)\Pr_A(C) + \Pr_A(\neg C \to B)\Pr_A(\neg C).$$

Using probabilistic export-import and The Thesis, we can expand this to an expression involving only the original probability distribution Pr:

$$\Pr(A)\Pr(A \to B) = \Pr(A \wedge C \to B)\Pr(A \wedge C) + \Pr(A \wedge \neg C \to B)\Pr(A \wedge \neg C).$$

Note that this expression follows simply from The Thesis and the laws of probability theory. The rule used in Lewis's argument, on the other hand, depends on further assumptions about how conditionals behave under probability operators and is not a consequence of the rules for probability.

Using this version of the law of total probability for conditionals, we can see that the triviality result does not go through. The triviality result followed from setting $C = B$ in the conditional law of total probability. If we do this, we get

$$\Pr(A)\Pr(A \to B) = \Pr(A \wedge B \to B)\Pr(A \wedge B) + \Pr(A \wedge \neg B \to B)\Pr(A \wedge \neg B) =$$

$$\Pr(A \wedge B \to B)\Pr(A \wedge B) + \Pr(A \wedge \neg B \to B)\Pr(A \wedge \neg B) =$$

$$1 * \Pr(A \wedge B) + 0 * \Pr(A \wedge \neg B) = \Pr(A \wedge B),$$

which is just the ratio definition of conditional probability, $\Pr(A \to B) = \frac{\Pr(A \wedge B)}{\Pr(A)}$.

This shows that Lewis's triviality result makes an assumption which is at odds with The Thesis. While The Thesis entails a natural way to calculate total probability for a conditional, Lewis uses an alternative calculation which does not follow from The Thesis or the laws of probability. This casts doubt on Lewis's assumption, which will be revisited as a consequence of the operator theory of conditionals in §4.

3 Bradley's Triviality Result

I have argued that Lewis's triviality result is at odds with The Thesis because of the version of the conditional law of total probability it relies on. It is natural to wonder whether a similar explanation can be given for other triviality results. One of the most popular such triviality arguments is that of Bradley

[10] Note that this expression, as well as a criticism of Lewis's argument along these lines, is also found in Egré and Cozic (2011).

(2000). Bradley argues that any probability distribution which satisfies the preservation condition is subject to triviality, where the preservation condition states that if $\Pr(A) > 0$ and $\Pr(B) = 0$, then $\Pr(A \to B) = 0$. The Thesis actually entails the Preservation Condition: if $\Pr(B) = 0$, then $\Pr(A \wedge B) = 0$, so $\Pr(B|A) = \frac{\Pr(A \wedge B)}{\Pr(A)} = 0$, provided $\Pr(A) > 0$.

Bradley's argument suggests that ascribing probabilities to conditionals allows us to contradict the preservation principle.[11] He assumes that we have a language with propositions A, B and $A \to B$ such that $A \not\models B$ and $A \to B \not\models B$. From this, he concludes that there must be a probability distribution Pr such that $\Pr(A) > 0$, $\Pr(A \to B) > 0$, and $\Pr(B) = 0$, violating the preservation principle. This follows from a standard view of logical consequence, where $X \models Y$ iff Y is true in every world where X is true, so if $X \not\models Y$, there must be a world where X is true and Y is not, and we can choose a Pr assigning positive probability to this world but zero probability to all Y worlds.

However, this argument again makes an assumption at odds with The Thesis. Even though $A \to B \not\models B$, every world which contributes probability to $A \to B$ is a world where both A and B are true, so it also contributes probability to B. To see this more directly, suppose $\Pr(A \to B) > 0$, where $\Pr(A \to B) = \Pr_A(B) = \sum_{\omega \in \Omega} \Pr_A(\omega)\chi_B(\omega)$. Since this sum is greater than 0, there is a world ω such that $\chi_B(\omega) = 1$ and $\Pr_A(\omega) > 0$. Since $\Pr_A(\omega) = \frac{\Pr(\omega)\chi_A(\omega)}{\Pr(A)}$ and $\Pr_A(\omega) > 0$, $\Pr(\omega) > 0$. Then since $\Pr(\omega) > 0$ and $\chi_B(\omega) = 1$, ω contributes probability to $\Pr(B) = \sum_{\omega \in \Omega} \Pr(\omega)\chi_B(\omega)$, so $\Pr(B) > 0$. Thus, we cannot conclude from the fact that $A \to B \not\models B$ that there is some Pr such that $\Pr(A \to B) > 0$ but $\Pr(B) = 0$, as required in Bradley's argument.

Thus, Bradley's argument assumes that, since $A \to B \not\models B$, there must be a world in Ω which we can assign probability to that contributes to $\Pr(A \to B)$, but not to $\Pr(B)$. If we accept The Thesis as formulated in the possible worlds framework, however, this is not the case. Like Lewis's result, Bradley's argument relies on an assumption at odds with The Thesis. In the next section, I argue that this is not a coincidence; both of these principles follow from the operator theory of conditionals, which is incompatible with The Thesis.

4 The Operator Theory and Triviality Results

I have argued that the triviality results of both Lewis and Bradley make assumptions which are not supported by The Thesis. These assumptions, however, follow from a semantic picture of conditionals which I call the operator theory. According to the operator theory, the conditional connective $\to$ joins two propositions, A and B, to form a new proposition $A \to B$. The operator theory assumes not just that $A \to B$ is a new proposition, but also that $A \to B$ embeds under other operators as a stand-alone proposition. The main

[11] My presentation here follows Khoo and Santorio (2018).

competing approach to the operator theory is the restrictor theory, which argues that the antecedent of the conditional serves to restrict other operators rather than to combine with the consequent to form a stand-alone proposition. The restrictor theory does not, however, entail that a bare conditional is not a proposition: for Kratzer (1986, 2012), for example, bare conditionals have a covert necessity modal 'must' which serves as the operator that the antecedent restricts, allowing for a propositional account of bare conditionals.

We can understand the difference between the two theories by considering the logical form of sentences of interest in this paper:

(1) The probability that if A, then B is α.

Here, the conditional expression 'If A, then B' is embedded under a probability operator, 'the probability of P is α', which applies to some proposition P.[12] According to the operator theory of the conditional, the conditional 'If A, then B' embeds as a proposition $P_{A\to B}$, so the logical form of (1) is:

(2) The probability that $P_{A\to B}$ is α.

On the restrictor theory of the conditional, however, the antecedent of the conditional functions to restrict the domain of a modal operator rather than to join with the consequent to form a proposition. In the case of (1), the modal operator is the probability operator, so the logical form is:

(3) The probability, given A, of B is α.

The difference between the operator and restrictor theories becomes clearer if we write down the probability operator more explicitly, following the discussion in §1. For a proposition B and a probability distribution over accessible worlds (Ω, Pr), the probability of B is α iff $\mathrm{Pr}(B) = \sum_{\omega\in\Omega} \mathrm{Pr}(\omega)\chi_B(\omega) = \alpha$; we can write this as $[\![\mathrm{Pr}(B) = \alpha]\!]^{w,\Omega} = 1$. Then, assigning a probability to conditionals, or interpreting the semantics of (1), requires specifying the truth conditions of $[\![\mathrm{Pr}(A \to B) = \alpha]\!]^{w,\Omega}$. On the operator theory, the conditional $A \to B$ is a proposition $P_{A\to B}$ with characteristic function $\chi_{A\to B}$, so $[\![\mathrm{Pr}(A \to B) = \alpha]\!]^{w,\Omega} = [\![\mathrm{Pr}(P_{A\to B}) = \alpha]\!]^{w,\Omega} = 1$ iff $\mathrm{Pr}(P_{A\to B}) = \sum_{\omega\in\Omega} \mathrm{Pr}(\omega)\chi_{A\to B}(\omega) = \alpha$. On this approach, the conditional embeds under probability operators as any other proposition does; the probability of the conditional is given by the expected value of the associated characteristic function.

On the restrictor theory, however, the conditional does not function as a stand-alone proposition; in fact, the language does not even have a proper conditional connective. Instead, we can treat the antecedent if-clause as adding an extra argument which restricts the domain of the probability operator. In this case, we can write the semantics of (1) as $[\![\mathrm{Pr}(\text{if } A)(B) = \alpha]\!]^{w,\Omega} =$

[12]For more on modal operators in general, see Kratzer (1981, 2012), and on probability operators specifically, see Yalcin (2007, 2010).

$[\![\Pr(B) = \alpha]\!]^{w,\Omega+A}$, where $\Omega+A$ is the probabilistic context (or modal domain) updated by A. We expect the updated probabilistic context to be restricted to those worlds where A is true, Ω_A, with the probability distribution Pr updated to the distribution $\Pr_A$.[13] Thus, $[\![\Pr(B) = \alpha]\!]^{w,\Omega+A} = [\![\Pr_A(B) = \alpha]\!]^{w,\Omega_A}$, which is equal to $[\![\Pr_A(B) = \alpha]\!]^{w,\Omega}$ since $\Pr_A$ is the same over both Ω and Ω_A. Therefore, on the restrictor theory, $[\![\Pr(\text{if } A)(B) = \alpha]\!]^{w,\Omega} = 1$ iff $\Pr_A(B) = \alpha$. Abusing notation, we can use the conditional symbol $\Pr(A \to B)$ as a placeholder for $\Pr(\text{if } A)(B)$; in this case, the probability of the conditional $\Pr(A \to B)$ is given by the conditional probability $\Pr_A(B)$. Thus, on this approach to probability operators and domain restriction, the restrictor theory entails The Thesis.

This shows that, on this approach to probability operators and domain restriction, the restrictor theory entails The Thesis. For the operator theory of the conditional, on the other hand, The Thesis is a substantial assumption which may be false. In fact, the triviality results show that the operator theory is incompatible with The Thesis. First, I will argue that the two problematic assumptions highlighted for Lewis's and Bradley's results follow from the operator theory. Then, using Hájek's (1989, 2012) 'Wallflower Argument' and following Egré and Cozic's undefinability argument, I will provide a more direct argument that the operator theory is inconsistent with The Thesis.

For Lewis's result, the problematic assumption is the conditional law of total probability, $\Pr(A \to B) = \Pr(A \to B|C)\Pr(C) + \Pr(A \to B|\neg C)\Pr(\neg C)$. On the operator theory, $A \to B$ embeds under probability operators as a proposition, so $\Pr(A \to B|C) = \Pr(P_{A\to B}|C)$. Using the definition of conditional probability, this becomes $\frac{\Pr(P_{A\to B}\wedge C)}{\Pr(C)}$, which is just $\frac{\Pr((A\to B)\wedge C)}{\Pr(C)}$. This consequence of the operator theory validates Lewis's conditional law of total probability:

$$\Pr(A \to B|C)\Pr(C) + \Pr(A \to B|\neg C)\Pr(\neg C) =$$

$$\frac{\Pr(C \wedge (A \to B))}{\Pr(C)}\Pr(C) + \frac{\Pr(\neg C \wedge (A \to B))}{\Pr(\neg C)}\Pr(\neg C) =$$

$$\Pr(C \wedge (A \to B)) + \Pr(\neg C \wedge (A \to B)) = \Pr(A \to B).$$

Thus, the operator theory of conditionals validates this assumption of Lewis's triviality argument.

Similarly, the operator theory confirms the principle needed for Bradley's triviality result: if $A \not\models B$ and $A \to B \not\models B$, then there must be a probability distribution Pr such that $\Pr(A) > 0$, $\Pr(A \to B) > 0$, and $\Pr(B) = 0$. To see that this is true, assume $A \not\models B$ and $A \to B \not\models B$. Then we can find worlds ω_1 and ω_2 such that $\chi_A(\omega_1) = 1$ and $\chi_{A\to B}(\omega_2) = 1$ but $\chi_B(\omega_1) = 0$ and $\chi_B(\omega_2) = 0$. Let π be a probability distribution assigning positive probability to both ω_1 and ω_2 and 0 to all other worlds. Then $\pi(A) > 0$ and $\pi(B) = 0$, and the operator theory specifies that $\pi(A \to B) = \sum_{\omega\in\Omega}\pi(\omega)\chi_{A\to B}(\omega) > 0$.

[13] An argument for why conditional probability is natural to use for the restrictor theory is given in Egré and Cozic (2011).

Thus, π is a distribution satisfying $\pi(A) > 0$, $\pi(A \to B) > 0$, and $\pi(B) = 0$, violating the preservation condition and leading to the conclusion of Bradley's triviality result.

These arguments show that the operator theory of conditionals validates the principles underlying Lewis's and Bradley's triviality results which conflict with The Thesis. However, it is worth noting that both Lewis's and Bradley's arguments make additional assumptions: Lewis assumes the probability assignments $\Pr(A \to B|B) = 1$ and $\Pr(A \to B|\neg B) = 0$ and Bradley assumes the preservation condition. Since I argued that these assumptions follow directly from The Thesis, this suggests that the operator theory and The Thesis are incompatible.

However, we can provide a more direct line of argument to show that the operator theory and The Thesis are incompatible, even without these additional assumptions; this is Hájek's (1989, 2012) 'Wallflower' argument. For a simple version of the argument, consider Hájek's example of a fair three-ticketed lottery. Since there is one winning ticket, there are three possible worlds $\omega_1, \omega_2, \omega_3$ corresponding to each of the three tickets winning, and each world has probability $\Pr(\omega_i) = 1/3$. According to The Thesis, the probability that 'If ticket 1 or 2 wins, then ticket 1 wins' is 1/2, but 1/2 is not attainable as the probability that some proposition is true; assuming the operator theory, the only possible probabilities for the conditional proposition are sums of $\Pr(\omega_i)$: 0, 1/3, 2/3, and 1. This shows that, if the operator theory is correct, there is some probability space $(\Omega, \Pr)$ such that The Thesis $\Pr(A \to B) = \Pr(B|A)$ is untenable on $(\Omega, \Pr)$, so The Thesis cannot hold universally. In fact, Hájek's argument extends further: for any finite $(\Omega, \Pr)$ with non-trivial Pr and $|\Omega| > 2$, there is a conditional probability value which cannot be attained as the probability of any proposition. Hájek's Wallflower argument shows that the operator theory is incompatible with The Thesis without any additional assumptions. Note, in addition, that neither The Thesis nor the restrictor theory impose the requirement that the probability of the conditional is the probability that a proposition is true, so Hájek's argument is not applicable without assuming the operator theory.

5 Restrictor Conditionals Without Triviality

I have argued that The Thesis only gives rise to triviality under the assumptions of the operator theory: the results of Lewis, Bradley, and Hájek all rely on assumptions which are at odds with The Thesis, but which follow from the operator theory. I have also argued that the restrictor theory is a competitor to the operator theory which entails The Thesis and, without the assumption of the operator theory, can avoid the triviality results. However, Charlow (2016) provides new interpretations of the triviality results intended to apply to the restrictor theory.

Charlow's results, however, do not succeed in establishing triviality for restrictor conditionals. This is because his arguments make use of assumptions from the operator theory which are at odds with the restrictor interpretation. Charlow follows Kratzer in interpreting a bare conditional as a restricted necessity modal, so the conditional 'If A, then B' is interpreted as the modal $must(A)(B)$, which requires that B must be the case when A is added to the body of information. Using $\rightarrow$ to represent the conditional and Ω to represent the body of information, this means that $[\![A \rightarrow B]\!]^{w,\Omega} = [\![must(A)(B)]\!]^{w,\Omega} = [\![must(B)]\!]^{w,\Omega+A}$, representing a standard interpretation of conditional semantics on the restrictor theory.

However, Charlow takes the restrictor theory to define a conditional proposition, or a set of possible worlds in which the conditional is true, and embeds the conditional under probability operators as if it were a stand-alone proposition. While the first assumption, that the conditional defines a proposition, is true on many restrictor theories, the second assumption utilizes the operator theory rather than the restrictor theory. Charlow makes use of the operator theory in his arguments by invoking the two principles identified as problematic in the arguments of Lewis and Bradley. In applying Lewis's triviality argument to restrictor conditionals, Charlow (2016, p. 547) assumes Lewis's version of the conditional law of total probability:

$$\Pr(must(A)(B)) = \Pr(must(A)(B)|B)\Pr(B) + \Pr(must(A)(B)|\neg B)\Pr(\neg B).$$

However, this implicitly makes the assumption that any propositional account of the conditional also requires that conditionals embed under probability operators as stand-alone propositions. Furthermore, this assumption is one which the restrictor theory of conditionals is designed to challenge: conditionals do not embed under modal operators as stand-alone propositions, with the antecedent instead serving to restrict the modal domain over which the consequent is evaluated. The same problem arises for Charlow's extension of Bradley's triviality result to restrictor conditionals, which relies on the condition identified as problematic in §3, which Charlow refers to as the Existence Lemma (p. 554).

Charlow could respond by claiming that a probability function need not always represent an overt probability operator. While such an overt operator is clearly present in a sentence like 'The probability that if A, then B is p,' it would not be present if the probability represents an agent's credence or subjective probability in the conditional 'If A, then B.' In this case, it may be natural to think that, even on the restrictor theory, the agent's credences must target the conditional proposition $must(A)(B)$. In this case, the conditional would embed in the probability function (no longer conceived of as a modal operator) as a proposition, allowing one to make use of the assumptions from Lewis's and Bradley's triviality results and to prove that triviality also applies to the restrictor semantic theory.[14] This seems to be Charlow's approach: 'if the conditional expresses a proposition, we may (indeed, should) flatly insist that the probability of that proposition must, e.g., obey the Law of Total

[14]This objection to the restrictor analysis of triviality is raised explicitly in Rothschild (2013).

Probability' (p. 538, n.6). This assumption is also perfectly consistent with the restrictor semantics: if there is no overt probability operator, the restrictor semantics does not specify how one must interpret a conditional within the scope of a credence or subjective probability function.

However, this becomes problematic if we adopt a more general restrictor principle: whenever a conditional is evaluated within a set of possible worlds, the antecedent restricts the domain of worlds rather than combining with the consequent to produce a stand-alone proposition. A consequence of this principle is that, whenever we evaluate the probability of a conditional 'If A, then B' within a probabilistic context (Ω, Pr), we do so by restricting the context by the antecedent A to $(\Omega_A, \mathrm{Pr}_A)$ and evaluating the probability of the consequent in this context. In the last section, I argued that the restrictor semantics for overt probability operators is a special case of this: the modal domain of the probability operator supplies the probabilistic context and the restrictor semantics tells us to evaluate the conditional by restricting this context.

If the probability function represents subjective probability rather than an overt modal, we can develop an analogous restrictor epistemology of conditionals from the restrictor principle. Generalizing both Stalnaker's (1984) model where an agent's beliefs are represented by a set of possible worlds and Bayesian epistemology where an agent's beliefs are represented by a credence function (Hartmann and Sprenger, 2010), we can represent an agent's beliefs by a probabilistic context (Ω, Pr). Then, when an agent evaluates his or her credence in a conditional 'If A, then B,' the restrictor principle tells us that the agent should restrict the probabilistic context to $(\Omega_A, \mathrm{Pr}_A)$ and evaluate B in this context. This is reminiscent of the suppositional approach to the conditional (Edgington, 1995), which can be summarized by Ramsey's suggestion that agents evaluate a conditional $p \to q$ by 'adding p hypothetically to their stock of knowledge and arguing on that basis about q' (Ramsey, 1990, p. 155). On this epistemology, applying a credence function to a conditional yields the same result as embedding a conditional under an overt probability operator: the probability of the conditional is not calculated as the likelihood that the conditional proposition is true, but rather as the likelihood that the consequent is true on the supposition of the antecedent. As before, this entails The Thesis and is at odds with the problematic assumptions of Lewis's and Bradley's triviality results, which follow from an operator epistemology of conditionals.

The restrictor principle is a plausible generalization of restrictor semantics and leads to a compelling and familiar epistemology for the probability of conditionals. Furthermore, such an alternative to the operator theory is necessary for The Thesis to hold without triviality. As soon as one introduces an operator interpretation of the conditional within a probabilistic context, any attempt to implement The Thesis results in triviality. This is what happens in Charlow's results: while he utilizes a restrictor semantics of the conditional, he assumes that the conditional interacts with the probabilistic context as an operator rather than as a restrictor. If one adopts the general restrictor principle, this assumption fails to hold and Charlow's triviality arguments no longer apply to the restrictor theory.

This shows that Charlow's arguments that the restrictor theory is subject to triviality rely on an operator interpretation of conditionals. While this operator interpretation is not explicitly ruled out by the restrictor semantics, it is likely an assumption many advocates of the restrictor theory would take issue with. This challenge also arises for other recent triviality results which may be thought to apply to the restrictor theory. Korzukhin (2014), for example, develops a triviality result aimed at context-sensitive conditionals which also relies on the version of conditional total probability which follows from the operator theory (p. 182). One can avoid these triviality results by adopting the general restrictor principle, where conditionals restrict any set of worlds they are evaluated within.

6 Conclusion

In this paper, I provided an interpretation of The Thesis that $\Pr(A \to B) = \Pr(B|A)$ in a possible worlds formalism and argued that the triviality results make assumptions which are at odds with The Thesis. I showed how these assumptions follow from the operator theory of conditionals and how rejecting the operator theory in favor of the restrictor theory can avoid the triviality results. Furthermore, I argued that recent attempts to attribute triviality to restrictor conditionals fail insofar as they make assumptions which follow from the operator theory. The plausibility of The Thesis and its relationship with the restrictor theory of conditionals provides evidence for abandoning the operator theory as a way of evaluating conditionals in a probabilistic context.

References

Adams, E. (1975), *The Logic of Conditionals: An Application of Probability to Deductive Logic*, Springer Science & Business Media.

Bacon, A. (2015), 'Stalnaker's thesis in context', *The Review of Symbolic Logic* **8**(1), 131–163.

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

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

Capinski, M. and Kopp, P. E. (2013), *Measure, integral and probability*, Springer Science & Business Media.

Charlow, N. (2016), 'Triviality for restrictor conditionals', *Noûs* **50**(3), 533–564.

Douven, I. and Verbrugge, S. (2010), 'The Adams family', *Cognition* **117**(3), 302–318.

Douven, I. and Verbrugge, S. (2013), 'The probabilities of conditionals revisited', *Cognitive Science* **37**(4), 711–730.

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

Egré, P. and Cozic, M. (2011), 'If-clauses and probability operators', *Topoi* **30**(1), 17–29.

Evans, J., Handley, S., Neilens, H. and Over, D. (2007), 'Thinking about conditionals: A study of individual differences', *Memory & Cognition* **35**(7), 1772–1784.

Fitelson, B. (2014), 'The strongest possible Lewisian triviality result', *Thought: A Journal of Philosophy* **4**(2), 69–74.

Hadjichristidis, C., Stevenson, R. J., Over, D. E., Sloman, S. A., Evans, J. S. B. and Feeney, A. (2001), On the evaluation of if p; then q conditionals, *in* 'Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society', Psychology Press, p. 381.

Hájek, A. (1989), 'Probabilities of conditionals: revisited', *Journal of Philosophical Logic* pp. 423–428.

Hájek, A. (2012), 'The fall of "Adams' thesis"?', *Journal of Logic, Language and Information* **21**(2), 145–161.

Hartmann, S. and Sprenger, J. (2010), 'Bayesian epistemology', *Routledge companion to epistemology* pp. 609–620.

Kaufmann, S. (2001), Aspects of the Meaning and Use of Conditionals, PhD thesis, Stanford University.

Khoo, J. (2016), 'Probabilities of conditionals in context', *Linguistics and Philosophy* **39**(1), 1–43.

Khoo, J. and Mandelkern, M. (2018), 'Triviality results and the relationship between logical and natural languages', *Mind* **128**(510), 485–526.

Khoo, J. and Santorio, P. (2018), 'Lecture notes: Probability of conditionals in modal semantics'.
URL: *http://www.justinkhoo.org/KS-NASSLLI2018.pdf*

Korzukhin, T. (2014), 'Contextualist theories of the indicative conditional and Stalnaker's thesis', *Thought: A Journal of Philosophy* **3**(3), 177–183.

Kratzer, A. (1981), The notional category of modality, *in* 'Words, Worlds and Contexts: New Approaches in World Semantics', Berlin: de Gruyter, pp. 38–74.

Kratzer, A. (1986), 'Conditionals', *Chicago Linguistics Society* **22**(2), 1–15.

Kratzer, A. (2012), *Modals and Conditionals: New and Revised Perspectives*, Vol. 36, Oxford University Press.

Lewis, D. (1976), Probabilities of conditionals and conditional probabilities, *in* 'Ifs', Springer, pp. 129–147.

Mandelkern, M. (2018), 'Import-export and 'and'', *Philosophy and Phenomenological Research* .
URL: *https://doi.org/10.1111/phpr.12513*

Oberauer, K., Geiger, S., Fischer, K. and Weidenfeld, A. (2007), 'Two meanings of "if"? individual differences in the interpretation of conditionals', *The Quarterly Journal of Experimental Psychology* **60**(6), 790–819.

Oberauer, K. and Wilhelm, O. (2003), 'The meaning(s) of conditionals: Conditional probabilities, mental models, and personal utilities.', *Journal of Experimental Psychology: Learning, Memory, and Cognition* **29**(4), 680–693.

Over, D. and Evans, J. (2003), 'The probability of conditionals: The psychological evidence', *Mind & language* **18**(4), 340–358.

Over, D., Hadjichristidis, C., Evans, J., Handley, S. and Sloman, S. (2007), 'The probability of causal conditionals', *Cognitive Psychology* **54**(1), 62–97.

Ramsey, F. (1990), General propositions and causality, *in* 'F. Ramsey, Philosophical Papers', pp. 145–163.

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

Rothschild, D. (2014), 'Capturing the relationship between conditionals and conditional probability with a trivalent semantics', *Journal of Applied Non-Classical Logics* **24**(1-2), 144–152.

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

Stalnaker, R. (1984), *Inquiry*, MIT Press, Cambridge, Mass.

Van Fraassen, B. (1976), Probabilities of conditionals, *in* 'Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science', Springer, pp. 261–308.

van Wijnbergen-Huitink, J., Elqayam, S. and Over, D. E. (2015), 'The probability of iterated conditionals', *Cognitive Science* **39**(4), 788–803.

Von Fintel, K. and Heim, I. (2011), 'Intensional semantics'.
URL: *http://web.mit.edu/fintel/fintel-heim-intensional.pdf*

Yalcin, S. (2007), 'Epistemic modals', *Mind* **116**(464), 983–1026.

Yalcin, S. (2010), 'Probability operators', *Philosophy Compass* **5**(11), 916–937.