Probability for epistemic modalities

Simon Goldstein and Paolo Santorio

June 28, 2021

Abstract

This paper develops an information sensitive theory of the semantics and probability of conditionals and statements involving epistemic modals. The theory validates a number of principles linking probability and modality, including the principle that the probability of a conditional *If A, then C* equals the probability of C, updated with A. The theory avoids so-called triviality results, which are standardly taken to show that principles of this sort cannot be validated. To achieve this, we deny that rational agents update their credences via conditionalization. We offer a new rule of update, Hyperconditionalization, which agrees with Conditionalization whenever nonmodal statements are at stake, but differs for modal and conditional sentences.

1 Introduction

Our topic is the semantics and epistemology of epistemic discourse, which includes indicative conditionals, like (1), and *must-* and *might-*claims, like (2) and (3)).

(1) If Frida danced, Maria danced.

(2) Frida must be dancing.

(3) Frida might be dancing.

The classical account of epistemic statements treats them as statements concerning worlds related to actuality via an appropriate accessibility relation (see Kratzer 2012, among others). Over the past decades, this account has come under attack from two independent directions. One tradition is motivated by nonstandard logical features of epistemic discourse, like the apparently inconsistency of *might* A and *not* A. Accounts in this tradition develop an information-sensitive semantics, on which meanings are characterized in terms of information states and update, rather than truth conditions.[1] A second tradition is motivated by the interaction between conditionals and probability. It is intuitive that one's credence in a conditional *If A, then C* should equal one's conditional credence $Pr(\mathsf{C} \mid \mathsf{A})$—a principle commonly referred to as 'Stalnaker's Thesis'. But a battery of triviality results show that Stalnaker's Thesis cannot be vindicated by a combination of classical truth-conditional semantics and classical Bayesian theories of credence. In light of this, theorists have developed non-truth-conditional accounts of conditionals aimed at vindicating probabilistic judgments.[2]

These two traditions have had little contact so far. This paper shows that they can and should be linked. We give an information-sensitive semantics and probability theory for epistemic discourse that not only captures informational inferences, but also vindicates a number of principles linking modality and probability. The semantics is a minimal variant of so-called path semantics, developed and defended in Santorio 2018, 2019b. This semantics is conservative with regard to existing accounts: it merely introduces extra structure on the well-known framework of information-sensitive semantics for epistemic modality. As a result, it integrates smoothly with current semantic theories, both compositionally and in terms of the resulting logic. We show that this semantics can be combined with a theory of probability and update that vindicates several intuitive bridge principles between modality and probability. In particular, using '$Pr_{\mathsf{A}}(\mathsf{B})$' to denote the probability of B, updated on A, we vindicate the following:

Update Thesis. For all rational Pr and all A, B:

$$Pr(\mathsf{A} \to \mathsf{B}) = Pr_{\mathsf{A}}(\mathsf{B})$$

In addition, we also vindicate a restricted version of Stalnaker's Thesis:

Stalnaker's Thesis^{-}. For all rational $Pr : Pr(\mathsf{A}) > 0$ and all descriptive A, B:

$$Pr(\mathsf{A} \to \mathsf{B}) = Pr(\mathsf{B} \mid \mathsf{A})$$

To achieve these goals, we develop a new theory of how rational agents update their credences over time. One of the lessons of triviality is that rational update of credences cannot be modeled, as Bayesians standardly do, via the rule of conditionalization. In place of conditionalization, we develop a new update rule, which we call 'Hyperconditionalization'. Hyperconditionalization agrees with conditionalization for ordinary sentences, but differs for epistemic sentences.

[1]There are several different kinds of information sensitive frameworks. For representative work within update semantics, see Veltman 1996, Groenendijk et al. 1996, Beaver 2001, Gillies 2004, and Willer 2013. For other information sensitive accounts, see Yalcin 2007, Swanson 2011, Klinedinst and Rothschild 2012, Swanson 2012, Moss 2015, Yalcin 2015, Ninan 2016, Mandelkern 2019, and Goldstein 2020. For useful overviews of the literature, see von Fintel and Gillies 2007 and Willer 2015.

[2]The first triviality result was presented in Lewis 1976; see also, among many, Hájek and Hall 1994, Bradley 2000, 2007, Charlow 2015, Goldstein 2017. As for no truth value accounts: the foundational text for all accounts in this tradition is Adams 1975; see also Edgington 1995. See van Fraassen 1976, Kaufmann 2009, 2015, Bacon 2015 for accounts that try to vindicate the Thesis.

Our results build on existing literature, and in particular on an influential line of thinking about conditionals and probability that traces back to van Fraassen 1976 and includes important work by Stefan Kaufmann (2009, 2015) and Richard Bradley (2000; 2007; 2012), among others. But our contribution is unique in a number of ways. (i) We offer a new diagnosis of triviality: conditionalization produces the wrong results when modal statements are involved. This diagnosis is backed by extensive theoretical and empirical arguments, in §3. (ii) Our semantics is compositionally integrated with existing systems for modals and conditionals. Our semantics is a refinement of standard information sensitive semantics. (iii) Existing accounts vindicate Stalnaker's Thesis at the price of a strong form of contextualism: the proposition expressed by a conditional changes whenever the epistemic state of the speaker changes. Conversely, our account doesn't rely on contextualism. One by-product of this is that, differently from proponents of contextualism, our version of Stalnaker's Thesis is robust under update: by rationally updating an epistemic state that vindicates the equivalence of $Pr(\mathsf{A} \to \mathsf{B})$ and $Pr(\mathsf{B} \mid \mathsf{A})$, we reach an epistemic state where the equivalence still holds (keeping the interpretation of the conditional unchanged).[3]

The paper is organized as follows. §§2–3 introduce triviality results and propose our diagnosis. §4 introduces our semantics. §5 presents an informal overview of our theory of probability. §6 develops this theory in detail, and §7 shows how update works. §8 addresses some further issues. §9 explores the extent to which our theory validates various strengthenings of Stalnaker's Thesis. The technical results are proved in appendices to the main body of the paper.

Before starting, a piece of terminology: throughout the paper, we say that a sentence is *descriptive* if it doesn't contain an epistemic modal or conditional.

2 Triviality

2.1 The issue of probabilities of conditionals

The philosophical literature on triviality begins with Stalnaker's Thesis (see Stalnaker 1970). Stalnaker's Thesis is the claim that, for all indicative conditionals $\mathsf{A} \to \mathsf{C}$, the probability of $\mathsf{A} \to \mathsf{C}$ equals the conditional probability $Pr(\mathsf{B} \mid \mathsf{A})$.

Stalnaker's Thesis (ST). For all rational $Pr : Pr(\mathsf{A}) > 0$ and all A, B:

$$Pr(\mathsf{A} \to \mathsf{B}) = Pr(\mathsf{B} \mid \mathsf{A})$$

There are many reasons to embrace ST. A first, simple argument is empirical. Ordinary judgments about probabilities of conditionals appear to align with it.[4] Suppose Maria might have tossed a fair die, and assess the probability of (4):

[3] See Norlin 2020; Norlin 2021 for recent work exploring the epistemological consequences of McGee 1985's semantics in a qualitative settings, which has some similar morals to our own.

[4] Though there are alleged local counterexamples to Stalnaker's Thesis; see Kaufmann 2004 for discussion.

(4) If Maria tossed the die, it landed on 1 or 2.

The natural answer is '1/3', which is also the conditional probability of the consequent, given the antecedent. These intuitions are confirmed by experimental evidence about probabilistic judgments about conditionals (see e.g. Evans and Over 2004, Cremers et al. 2017.)

A second, more general argument for ST links conditionals in natural language with the attitude of supposition. It is natural to think that evaluating a conditional (for either truth or probability) is equivalent to supposing the antecedent, and evaluating the consequent under that supposition. For example, assessing the probability of (4) appears to be equivalent to assessing the probability of the die landing 1 or 2, under the supposition that Maria did indeed toss the die. If, following standard Bayesian assumptions, we construe probabilities under supposition as conditional probabilities, we obtain ST.

The suppositional approach has a long pedigree in philosophy. Its central idea can be discerned in Ramsey 1926, and is developed formally and philosophically in Ernest Adams's important work. (see Adams 1975, as well as Adams 1998).[5] Among other things, the core idea of the suppositional approach promises to bridge in an interesting way semantics and epistemology. On the suppositional view, the semantic operation performed by *if*-clauses is an object language implementation of the operation of update of an epistemic state. Hence, if the suppositional theory is correct, there are interesting links between the semantics of natural language and a theory of supposition and update.

Unfortunately, ST has proven hard to vindicate. Starting with Lewis 1976, theorists have produced a vast array of "triviality" results showing that ST, in combination with seemingly safe assumptions about probability and conditionals, leads to unacceptable consequences. (Besides the original result in Lewis 1976, see Hájek and Hall 1994, Bradley 2000, 2007 among many.) In the remainder of this section, we present a simple triviality result for conditionals.

2.2 Triviality for conditionals

We start by outlining some basic assumptions. As we point out in §3, these assumptions are shared by a series of triviality results that go beyond conditionals.[6] Where Pr_{A} is the result of rationally updating credence function Pr on A (whatever the rational update procedure is):

[5] See, among others, Edgington 1995, Williams 2010 for more recent discussion.

[6] The characterization of these assumptions follows, in part Santorio and Williams.

TRIVIALITY QUINTET	
Identity.	For all Pr and all consistent A, $Pr_{\mathsf{A}}(\mathsf{A}) = 1$
Conjunction.	For all rational Pr, and for all A and B: $Pr(\mathsf{A} \wedge \mathsf{B}) \leq Pr(\mathsf{A})$
Ratio.	For all A and B: $Pr(\mathsf{B} \mid \mathsf{A}) = Pr(\mathsf{A} \wedge \mathsf{B})/Pr(\mathsf{A})$
Conditionalization.	For all rational Pr and all A: $Pr_{\mathsf{A}}(\bullet) = Pr(\bullet \mid \mathsf{A})$
Closure.	If Pr is rational, then $Pr_{\mathsf{A}}(\bullet)$ is rational, for any A such that $Pr(\mathsf{A}) > 0$.

Identity says that after updating on A, a subject should be certain of A. **Conjunction** says that the probability of a conjunction should be the lower bound of the probability of the conjuncts. **Ratio** defines conditional probabilities by the usual ratio of unconditional probabilities. **Closure** states that if we start from a rational credence distribution and update rationally on any claim, we reach a rational credence distribution. Finally, **Conditionalization** identifies the probabilities we reach via rational update (what we call *updated probabilities*) with conditional probabilities as defined by the ratio formula.

The triviality result we prove is a variant of a triviality result put forward by Bradley (2000, 2007). First, we observe that, given minimal principles **ST**, entails the following principle:[7]

> **Positive Preservation (PP).**
> For all rational Pr, and for all descriptive A, B such that $Pr(\mathsf{A}) > 0$:
> If $Pr(\mathsf{B}) = 1$, then $Pr(\mathsf{A} \rightarrow \mathsf{B}) = 1$

Positive Preservation also appears plausible on empirical grounds. If you have full credence that the die landed on 1 or 2, then you should also have full credence that, if Maria was the one to toss it, it landed on 1 or 2.

Now, given **Ratio**, **Closure**, and **Conjunction**, we can use **Positive Preservation** to prove that $Pr(\mathsf{A} \rightarrow \mathsf{B}) \geq Pr(\mathsf{B})$, on the assumption that $Pr(\mathsf{A}) > 0$.

i. $Pr(\mathsf{A} \rightarrow \mathsf{B}) \geq Pr((\mathsf{A} \rightarrow \mathsf{B}) \wedge \mathsf{B})$ (**Conjunction**)

ii. $Pr(\mathsf{A} \rightarrow \mathsf{B}) \geq Pr(\mathsf{B}) \times Pr(\mathsf{A} \rightarrow \mathsf{B} \mid \mathsf{B})$ (i, **Ratio**)

iii. $Pr(\mathsf{A} \rightarrow \mathsf{B}) \geq Pr(\mathsf{B}) \times Pr_{\mathsf{B}}(\mathsf{A} \rightarrow \mathsf{B})$ (ii, **Conditionalization**)

iv. $Pr_{\mathsf{B}}(\mathsf{B}) = 1$ (**Identity**)

[7] Given **Ratio**, and given the classical Bayesian assumption that $Pr(\mathsf{A} \wedge \mathsf{B}) = Pr(\mathsf{B})$ when $Pr(\mathsf{B}) = 1$, we have: if $Pr(\mathsf{B}) = 1$, then we have $Pr(\mathsf{B} \mid \mathsf{A}) = 1$. **Positive Preservation** immediately follows from here, given **ST**.

v. $Pr_{\mathsf{B}}(\mathsf{A} \to \mathsf{B}) = 1$ (iv, **PP**, **Closure**)

vi. $Pr(\mathsf{A} \to \mathsf{B}) \geq Pr(\mathsf{B}) \times 1 = Pr(\mathsf{B})$ (iii, v)

But this result is very implausible. Suppose that Maria tossed a fair die, and consider:

(5) The die landed on a prime.

(6) If the die landed on an even number, it landed on a prime.

Intuition suggests that $Pr(5) = 1/2$ (since 2, 3, and 5 are prime) and $Pr(6) = 1/3$. But the proof appears to establish that this is an irrational credal assignment, since it implies $Pr(6) \geq Pr(5)$.

This concludes our quick review of triviality for conditionals. Let us now turn to sketching our strategy for solving the problem.

3 Three notions of suppositional probability

This paper puts forward a theory of conditionals, probability, and update that vindicates the link between conditionals and probability outlined in §2, while avoiding triviality. The theory has a formal component, but the key idea can be stated and motivated informally. This section is dedicated to this task.

It's helpful to put on the table three notions, all of which have some claim to being (or being related to) a concept of 'suppositional probability'.

$Pr(\mathsf{A} \to \mathsf{C})$	probability of a conditional $\mathsf{A} > \mathsf{C}$
$Pr_{\mathsf{A}}(\mathsf{C})$	probability of C, updated on A
$Pr(\mathsf{C} \mid \mathsf{A})$	conditional probability of C, given A

The debate on probabilities of conditionals has focused on the viability of Stalnaker's Thesis, repeated below. Stalnaker's Thesis states the equivalence of the first and the third of these notions.

Stalnaker's Thesis (ST). For any A and B such that $Pr(\mathsf{A}) > 0$:

$$Pr(\mathsf{A} \to \mathsf{B}) = Pr(\mathsf{B} \mid \mathsf{A})$$

Throughout the debate, the notion of conditional probability has been invariably construed along classical Bayesian lines. This construal involves two elements. First, conditional probabilities are taken to conform to the ratio formula. I.e., as **Ratio** in §2 stated, $Pr(\mathsf{B} \mid \mathsf{A})$ is identified with the quotient $Pr(\mathsf{A} \wedge \mathsf{B})/Pr(\mathsf{A})$ (at least whenever $Pr(\mathsf{A}) > 0$).[8] Second, conditional probabilities so understood play an important role in update. In particular, the probability of B after learning A, i.e. $Pr_{\mathsf{A}}(\mathsf{B})$ is supposed to equal just the conditional probability $Pr(\mathsf{B} \mid \mathsf{A})$.

[8]This holds both if we take the ratio formula to provide a *definition* of conditional probability, and if we take conditional probabilities as primitive and assume that the ratio formula specifies a necessary equivalence that holds whenever the ratio of $Pr(\mathsf{A} \wedge \mathsf{B})/Pr(\mathsf{A})$ is defined (as recommended by Hájek 2003).

Conditionalization. For all rational Pr and all descriptive A, B:

$$Pr_{\mathsf{A}}(\mathsf{B}) = Pr(\mathsf{B} \mid \mathsf{A})$$

The combination of **Stalnaker's Thesis** and **Conditionalization** yields that all three notions related to the idea of suppositional probability coincide.

$$Pr(\mathsf{A} \to \mathsf{B}) \quad = \quad Pr_{\mathsf{A}}(\mathsf{B}) \quad = \quad Pr(\mathsf{B} \mid \mathsf{A})$$

As we pointed out in §2, an account that manages to vindicate the equivalence of these three notions is something like a theorist's dream. Unfortunately, the lesson of triviality result is just that these equivalences cannot all be vindicated—at least not without qualification.

The standard response, starting from Lewis 1976, has been to preserve the classical identification of conditional probabilities and updated probabilities, and to deny that the latter two equal probabilities of conditionals.

$$Pr(\mathsf{A} \to \mathsf{B}) \quad \neq \quad Pr_{\mathsf{A}}(\mathsf{B}) \quad = \quad Pr(\mathsf{B} \mid \mathsf{A})$$

This is a mistake. The right response is to hold on to the identification of probabilities of conditionals with updated probabilities, while denying that, in general, the latter two are properly captured by conditional probabilities as defined by the ratio formula.

$$Pr(\mathsf{A} \to \mathsf{B}) \quad = \quad Pr_{\mathsf{A}}(\mathsf{B}) \quad \neq \quad Pr(\mathsf{B} \mid \mathsf{A})$$

Since we reject that conditional probabilities, as classically construed, equal either probabilities of conditionals or updated probabilities, on our view both **Stalnaker's Thesis** and **Conditionalization**, in their unrestricted form, fail. Conversely, we vindicate the following:

Update Thesis. For all rational Pr and all A, B:

$$Pr(\mathsf{A} \to \mathsf{B}) = Pr_{\mathsf{A}}(\mathsf{B})$$

The Update Thesis says that the probability of any conditional $\mathsf{A} \to \mathsf{C}$ equals the probability of the consequent C, updated on the antecedent A. The Update Thesis is significant from a theoretical point of view. Via the Update Thesis, we can vindicate the idea that conditionals are object language devices for representing rational update via supposition.

At the same time, we don't reject the link between conditional probabilities and the other two notions of suppositional probability altogether. Our departure from the orthodoxy is restricted to special cases. Updated probabilities differ from conditional probabilities only for sentences that involve modal and conditional operators. But $Pr_{\mathsf{A}}(\mathsf{B})$ and $Pr(\mathsf{A} > \mathsf{B})$ equal $Pr(\mathsf{B} \mid \mathsf{A})$ whenever descriptive sentences—i.e., sentences that don't involve modal or conditional contents—are involved. As a result, our theory will also vindicate restricted versions of **Stalnaker's Thesis** and **Conditionalization**.

Stalnaker's Thesis$^-$. For all rational $Pr : Pr(\mathsf{A}) > 0$ and all descriptive A, B:

$$Pr(\mathsf{A} \to \mathsf{B}) = Pr(\mathsf{B} \mid \mathsf{A})$$

Conditionalization$^-$. For all rational Pr and all descriptive A, B:

$$Pr_{\mathsf{A}}(\mathsf{B}) = Pr(\mathsf{B} \mid \mathsf{A})$$

Hence, on our view, standard Bayesian tenets apply to the nonmodal and nonconditional fragment of our language. Given that, on our view, modal and conditional sentences do not express standard propositions and have a nonclassical logic, it is unsurprising that classical Bayesian tenets fail. A nonclassical semantics and logic requires a nonclassical theory of update.

Even with this qualification, one might worry that rejecting **Conditionalization** is the wrong reaction to triviality. In the rest of this section, we give two arguments to support our move: (i) triviality concerns not just conditonals, but a wide variety of expressions in language; (ii) there are straightforward counterexamples to the unrestricted versions of **Conditionalization** and **Stalnaker's Thesis**.

3.1 Argument #1: the generality of triviality

The problem of triviality was initially raised for **Stalnaker's Thesis**. But subsequent literature has shown that the focus on conditionals is misplaced: triviality is a problem affecting the connection between probability and epistemic modality in general.[9] On the one hand, we can derive triviality for conditionals without appealing to **ST**, but rather building on less controversial principles. On the other hand, triviality results can be derived also for modal statements that don't involve conditionals. Hence merely rejecting **ST** is insufficient to solve the problem.

In the rest of this section, we show how we can derive triviality results for necessity and possibility modals, using the same assumptions we used in §2:

[9] In fact, triviality arguably extends beyond epistemic modality. See Santorio and Williams for a triviality result concerning determinacy operators, and Williams 2012, a.o., for triviality results concerning counterfactuals.

TRIVIALITY QUINTET	
Identity.	For all Pr and all consistent A, $Pr_{\mathsf{A}}(\mathsf{A}) = 1$
Conjunction.	For all rational Pr, and for all A and B: $Pr(\mathsf{A} \wedge \mathsf{B}) \leq Pr(\mathsf{A})$
Ratio.	For all A and B: $Pr(\mathsf{B} \mid \mathsf{A}) = Pr(\mathsf{A} \wedge \mathsf{B})/Pr(\mathsf{A})$
Conditionalization.	For all rational Pr and all A: $Pr_{\mathsf{A}}(\bullet) = Pr(\bullet \mid \mathsf{A})$
Closure.	If Pr is rational, then $Pr_{\mathsf{A}}(\bullet)$ is rational, for any A such that $Pr(\mathsf{A}) > 0$.

The results of this section build on existing work, though our way of presenting and organizing the material is new.[10]

Necessity modals. We start with epistemic *must*, which we represent as '$\Box$'. Rather than PP, we assume:

> **Must Preservation (MP).** For all rational Pr and all A: If $Pr(\mathsf{A}) = 1$, then $Pr(\Box\mathsf{A}) = 1$

The empirical case for **Must preservation** mirrors the empirical case for an analogous principle in the logic of epistemic modality. Veltman 1985 and Yalcin 2007 argue that the following is a semantically valid argument:

(7) a. The house is empty.
  b. Therefore, the house must be empty.

(7) is an instance of what has become known as 'Lukasiewicz's principle'.

(8) **Lukasiewicz's Principle.** $\mathsf{A} \models \Box\mathsf{A}$

Here we remain neutral about the validity of Łukasiewicz's Principle. What matters to us is that the inference from (7-a) to (7-b) is *certainty-preserving*: if someone is certain that the house is empty, then they should also be certain that the house must be empty.

Now, assume **MP** and the principles in TRIVIALITY QUINTET. The proof is analogous to the one for conditionals:

i. $Pr(\Box\mathsf{A}) \geq Pr(\Box\mathsf{A} \wedge \mathsf{A})$ (**Conjunction**)

ii. $Pr(\Box\mathsf{A}) \geq Pr(\Box\mathsf{A} \mid \mathsf{A}) \times Pr(\mathsf{A})$ (i, **Ratio**)

[10] For results about necessity and possibility modals, see, among others, Schulz 2010, Charlow 2015, Russell and Hawthorne 2016, Beddor and Goldstein 2018, and Santorio 2019a.

iii. $Pr(\Box A) \geq Pr_A(\Box A) \times Pr(A)$ (iii, **Conditionalization**)

iv. $Pr_A(A) = 1$ (**Identity**)

v. $Pr_A(\Box A) = 1$ (iv, **MP**, **Closure**)

vi. $Pr(\Box A) \geq 1 \times Pr(A) = Pr(A)$ (iii, v)

The proof establishes that $Pr(\Box A) \geq Pr(A)$, i.e. that one's credence in *must* A is an upper bound for one's credence in A. But this seems absurd. Clearly one can have high credence in A, while having zero or near-zero credence in *must* A. To see this, consider a case discussed by Beddor and Goldstein 2018. Suppose Ari the burglar has been casing a house for hours. As far as she can tell, not a mouse is stirring. Consequently, Ari has high credence that the house is empty. Still, Ari, being an experienced burglar, grants that there is some possibility that an inconspicuous resident is inside. So she has low credence that the house must be empty. This combination of credences appears to be rational, contrary to the proof we just ran.

Possibility modals. For possibility modals like *might*, we also start from a specific principle about credences in *might*-claims. Where $\bar{A}$ denotes the negation of A:

> **Might Contradiction (MC).** Where Pr is a rational probability function:
> If $Pr(\bar{A}) = 1$, then $Pr(\Diamond A) = 0$.

Might Contradiction follows from **Must Preservation** given two orthodox assumptions: (i) if $Pr(A) = 1$, $Pr(\bar{A}) = 0$; (ii) *must* and *might* are duals, i.e. $\Box A =_{def} \neg\Diamond\neg A$.

There are intuitions backing **MC**. For an example: after an utterance of (9-a) is accepted, even asking the relevant *might*-claim, as in (9-b), seems out of place.

(9) a. A: It's not raining
b. B: # Might it be raining?

In addition to **MC**, we need to strengthen one of the principles in TRIVIALITY QUINTET. Rather than **Conjunction**, we assume **Total probability**:

> **Total probability.** For all A and B, $Pr(A) = Pr(A \wedge B) + Pr(A \wedge \neg B)$

In addition, we make use of two assumptions that seem uncontroversial: (i) A entails $\Diamond$A; (ii) if A entails B, $Pr_A(B) = 1$. We refer to the conjunction of these two assumptions below as **Entailment**.

With these principles in place, here is the triviality proof:

i. $Pr(\Diamond A) = Pr(\Diamond A \wedge A) + Pr(\Diamond A \wedge \bar{A})$ (**Total Probability**)

ii. $Pr(\Diamond A) = Pr(\Diamond A \mid A) \times Pr(A) + Pr(\Diamond A \mid \bar{A}) \times Pr(\bar{A})$ (i, **Ratio**)

iii. $Pr(\Diamond A) = Pr_A(\Diamond A) \times Pr(A) + Pr_{\bar{A}}(\Diamond A) \times Pr(\bar{A})$ (ii, Def of Pr_A)

iv. $Pr_A(A) = 1$ (**Identity**)

v. $Pr_A(\Diamond A) = 1$ and $Pr_{\bar{A}}(\Diamond A) = 0$ (iv, **Entailment**, **MC**, **Closure**)

vi. $Pr(\Diamond \mathsf{A}) = 1 \times Pr(\mathsf{A}) + 0 \times Pr(\bar{\mathsf{A}})$ (iii, v)

vii. $Pr(\Diamond \mathsf{A}) = Pr(\mathsf{A})$ (vi)

This result is, again, absurd. Suppose that a fair coin was flipped and I am fully ignorant about the outcome. I will have high credence in the claim that the coin might have landed heads, but only .5 credence in the claim that the coin landed heads. Yet this simple combination of attitudes is ruled out by our result.

To summarize: the assumptions in TRIVIALITY QUINTET can be used, in combination with various principles about modals and conditionals, to generate a battery of triviality results for all sentences involving epistemic modalities.

It's natural to think that the way to block all these results is to reject one of these assumptions. But, once we take this path, **Conditionalization** is the only reasonable choice. **Identity** and **Conjunction** seem too basic to question. Within our setup, **Ratio** is simply a definition: it defines conditional probabilities as ratio probabilities.[11] Similarly, **Closure** merely establishes that, by starting from a rational probability distribution and by performing rational update—whatever procedure rational update consists in—we reach another rational probability distribution. This much seems hard to give up. Conversely, **Conditionalization** makes a substantial claim: rationally updated probabilities are conditional probabilities, as they are defined by the ratio formula. This is this claim that we reject.

One might worry that, by rejecting **Conditionalization**, we are changing the subject from classical discussions of the conditionals-probability link. The notion of conditional probability, the thought goes, is crucially linked to the notion of update; the main theoretical role of conditional probabilities is just to be used in update. So, by changing the notion of update, we appear to be changing the notion of conditional probability as well. The response is that our notion of updated probability is genuinely linked to the classical notion of conditional probability. As we pointed out, we vindicate the restricted principle **Conditionalization**$^-$: updated probabilities *are* identical to conditional probabilities whenever modals and conditionals are not involved. So we are not swapping the standard concept of conditional probabilities for another one. We are simply amending it to cover cases that were never in its originally intended coverage.

3.2 Argument #2: failures of ST and Conditionalization

Our second argument is that there is empirical evidence that the unrestricted versions of **Conditionalization** and **Stalnaker's Thesis** fail when sentences involving conditionals and modals are involved. Moreover, these failures appear to be connected,

[11] An influential line of thinking about conditional probability (see e.g. Hájek 2003) rejects the idea that conditional probabilities should be defined via **Ratio**, and instead treat **Ratio** as a necessary truth that holds whenever the quotient of $Pr(\mathsf{A} \wedge \mathsf{B})/Pr(\mathsf{A})$ is defined. On this construal, **Ratio** is the assumption that fails on our theory. This doesn't make a difference for out key claim: updated probabilities fail to be equal to conditional probabilities, *understood as conforming to the ratio formula*, when modals and conditionals are involved.

since they happen in the same scenarios. These failures have gone unnoticed so far, so we discuss them in detail here.

Failure of Stalnaker's Thesis. Consider the following scenario:

> **Die.** Sarah tossed a fair, six-sided die; we have no information about the outcome.

Notice first that (10) should get probability 1, or near-1.[12]

(10) If the die landed even, then if it didn't land on two or four, it landed on six.

Let's now consider the conditional probability of the consequent given the antecedent of (10). Notice first that (11) and (12) should get, respectively, probability 1/4 (assuming Stalnaker's Thesis) and 1/2.

(11) If the die did not land on two or four, it landed on six.

(12) The die landed even.

Notice also that (12) is equivalent to the material conditional corresponding to (11). Hence, given the widely accepted assumption that indicative conditionals entail the corresponding material conditionals[13], (11) entails (12). As a result, the conditional probability of (11) given (12) is the following:

$$Pr(\neg(2\text{or}4) > 6 \mid \text{even}) = \frac{Pr(\neg(2 \text{ or } 4) > 6 \wedge \text{even})}{Pr(\text{even})} = \frac{Pr(\neg(2 \text{ or } 4) > 6)}{Pr(\text{even})} = \frac{1/4}{1/2} = \frac{1}{2}$$

So we have:

$$1 = Pr(\text{even} > (\neg(2 \text{ or } 4) > 6)) \neq Pr(\neg(2 \text{ or } 4) > 6 \mid \text{even}) = \frac{1}{2}$$

This is a counterexample to the unrestricted version of Stalnaker's Thesis.

Failure of Conditionalization. The same scenario that we used to produce a counterexample to **Stalnaker's Thesis** can be used to produce a counterexample to unrestricted Conditionalization. This supports our point that the two theses are related, and that they should be rejected together. Let us focus now just on (11) and (12), repeated below:

(11) If the die did not land on two or four, it landed on six.

(12) The die landed even.

[12] Depending on whether you are fine with assigning contingent propositions probability 1; for discussion, see, among many, Easwaran 2014.

[13] This principle is vindicated by the so-called Weak Centering principle: $\mathsf{A} > \mathsf{C} \vDash \mathsf{A} \supset \mathsf{C}$. There is evidence that Weak Centering doesn't hold in general (see among many McGee 1985, Khoo 2013), but it is uncontroversial for simple conditionals like (11).

As we saw, in the situation described one should assign, respectively, probability 1/4 and 1/2 to (11) and (12).

Suppose now that we learn (12) with certainty. What probability should we assign to (11), in the updated belief state? The natural answer is '1', or 'near-1'. Once we are certain that the die landed even, we should also take the conditional in (11) to be certain.

The relevant probabilities before and after the learning event are specified in this table:

	before update with 'even'	*after update with 'even'*
even	1/2	1
if not (two or four), six	1/4	1

Now, these probabilities are incompatible with the claim that learning takes place via **Conditionalization**. The following condition is a basic feature of conditional probabilities:

For any A, B: if $Pr(\mathsf{A}) < Pr(\mathsf{B})$, then $Pr(\mathsf{A} \mid \mathsf{B}) < 1$.

In fact, given the assumption that (11) entails (12), we obtain, via a calculation analogous to the one above, that our degree of belief in the conditional, conditional on the claim that the die landed even, should be exactly 1/2.

Hence the probability of (11), updated on (12), is different from the conditional probability of (11), given (12).

$$1 = Pr_{\text{even}}(\neg\, (2 \text{ OR } 4) > 6)) \neq Pr(\neg\, (2 \text{ OR } 4) > 6 \mid \text{even}) = \frac{1}{2}$$

This is a counterexample to the unrestricted version of **Conditionalization**.

3.3 Noncontextualism

Before stating our account, it's useful to contrast it with a different strategy for tackling triviality. Several writers (see e.g. van Fraassen 1976; Kaufmann 2009; Bacon 2015) appeal to a contextualist semantics to block triviality. The key idea is that conditionals express a different proposition when evaluated with respect to different epistemic states. This is particularly relevant when conditionals appear as arguments of probability functions that represent credences, since a change in credence distribution corresponds to a change in epistemic states. As a result, triviality proofs of the kind that we have presented in §2 suffer from systematic equivocation. The sentence '$\mathsf{A} \to \mathsf{C}$' expresses different proposition in $Pr(\mathsf{A} \to \mathsf{C})$ and in $Pr(\mathsf{A} \to \mathsf{C} \mid \mathsf{C})$.

The contextualist strategy, appropriately supplemented with a package of semantics and probability theory that assigns probabilities to conditionals, provides a response to triviality. Moreover, the strategy can be easily generalized to epistemic modals, thus providing a general solution to the family of triviality results reviewed in §2.

Despite these advantages, we don't pursue a contextualist solution. The reason is that we the kidn of contextualism needed to block triviality is implausible, and that an alternative noncontextualist solution is available.

In this paper, we don't have the space to offer a detailed critique of contextualism, but let us gesture towards a point of disagreement. Consider again the die scenario, and in particular (11):

(11) If the die did not land on two or four, it landed on six.

According to the contextualist, the conditional in (11) expresses different propositions before and after you learn that the die landed even. Before the learning event, it describes an epistemic state in which all outcomes for the die toss are open possibilities. After the learning event, it describes an epistemic states where only even numbers are open possibilities. More in general, (11) will express a different proposition whenever the speaker's epistemic state changes.

This extreme kind of context dependence, which appears to be by and large invisible to speakers[14], is not well-motivated. It is also not necessary. This paper shows how we can get a theory of the probabilities of epistemic modality and conditionals that avoids this kind of extreme contextualism.

To highlight our disagreement with the contextualist, let us close by pointing to a principle that we vindicate and that fails on contextualist accounts. On our view, the restricted version of Stalnaker's Thesis is robust under updating. If a subject's initial probability distribution vindicates **Stalnaker's Thesis**$^{-}$, and if they update rationally, they reach a probability distribution that also vindicates **Stalnaker's Thesis**$^{-}$—holding constant the content assigned to the conditional.

Robust Stalnaker (RS). For all rational Pr, all descriptive A, B, and all C:

$$Pr_{\mathsf{C}}(\mathsf{A} \to \mathsf{B}) = Pr_{\mathsf{C}}(\mathsf{B} \mid \mathsf{A})$$

Notice that **RS** is validated in the die scenario. Before the learning event, the intuitively correct degree of rational credence in (11) is 1/4, which corresponds to the conditional probability of EVEN, given $\neg$(2 OR 4). After the learning event, the intuitive rational credence for (11) is 1, which corresponds to the updated conditional probability.

[14]Why is it invisible to speakers? Suppose that Alice and Bob, who have different epistemic state, disagree on what credence one should assign to the proposition they each express via (i).

(i) If it's raining in Melbourne, it's raining in Sydney too.

According to the contextualist, Alice and Bob are talking past each other, since they express different propositions. For an analogy, consider a case where Alice and Bob disagree about what credence to assign to the proposition they each express by (ii)—when they consider (ii) at different places/times.

(ii) It is raining.

Clearly, in this second case Alice and Bob's disagreement is the result of a misunderstanding. But the case of the disagreement about (i) seems very different.

Conversely, contextualist accounts do not vindicate **RS**. For example, for the contextualist, the proposition originally expressed by (11) receives probability 1/2 after update. To our knowledge, **RS** is not vindicated by any existing account of conditionals and probability. Yet it is a natural desideratum, and one that seems obvious from a naïve look at the evidence.

3.4 The plan

The project of this paper is to develop a theory of conditionals, probability, and update that vindicates three main claims:

i. the **Update Thesis**;

ii. the restricted verson of **Stalnaker's Thesis**, i.e. **Stalnaker's Thesis**$^-$;

iii. the thesis that (ii) is robust under update, i.e. **Robust Stalnaker**.

Notice that, via (i) and (ii), it immediately follows that our theory also vindicates **Conditionalization**$^-$. Via the **Update Thesis**, $Pr(\mathsf{A} \to \mathsf{B}) = Pr_{\mathsf{B}}(\mathsf{A})$, for all A and B. Via **Stalnaker's Thesis**$^-$, $Pr(\mathsf{A} \to \mathsf{B}) = Pr(\mathsf{B} \mid \mathsf{A})$, for all descriptive A and B. It follows that $Pr_{\mathsf{A}}(\mathsf{B}) = Pr(\mathsf{B} \mid \mathsf{A})$, for all descriptive A and B.

In addition, the resulting theory will also vindicate the principles about credences in epistemic modal claims mentioned in this section. Hence we are going to give a general theory of the links between probability and epistemic modal language, that vindicates all the main intuitive bridge principles discussed so far.

4 Path semantics

Santorio 2019b develops an information sensitive semantics for modals and conditionals called 'path semantics'. We are going to adopt path semantics, and define notions of probability and update based on it.

We choose path semantics for three reasons. First, path semantics is an informational semantics, in the sense that it vindicates inferences that are typical of informational systems in the style of Veltman 1996 and Yalcin 2007. This will help us capture several of the principles in §3.

Second, path semantics, differently from all other informational semantics, vindicates Conditional Excluded Middle:

(13) **Conditional Excluded Middle (CEM).** $\models \mathsf{A} \to \mathsf{B} \vee \mathsf{A} \to \neg\mathsf{B}$

CEM is related to Stalnaker's Thesis, and so validating the former is a step towards validating the latter.[15]

[15] Suppose CEM is invalid. Then an agent can assign a credence of less than 1 to $\mathsf{A} \to \mathsf{B} \vee \mathsf{A} \to \neg\mathsf{B}$. But Stalnaker's Thesis implies that $Pr(\mathsf{A} \to \mathsf{B} \vee \mathsf{A} \to \neg\mathsf{B}) \geq Pr(\mathsf{A} \to \mathsf{B}) + Pr(\mathsf{A} \to \neg\mathsf{B}) = Pr(\mathsf{B} \mid \mathsf{A}) + Pr(\neg\mathsf{B} \mid \mathsf{A}) = 1$.

Finally, path semantics is a compositional semantics for natural language. In particular, it handles together modalized and conditional statements. This gives it an advantage other semantics that have been proposed to vindicate Stalnaker's Thesis, and which focus exclusively on the conditional case.[16]

In an informational semantics, modal sentences like $\Box$A or $\Diamond$A quantify over possible worlds in an information state. The central idea behind path semantics is that, differently from modal sentences, conditionals don't quantify over possible worlds. Rather, conditionals are selectional: they select a single world to use as the world of evaluation for the consequent.[17] The selectional idea is implemented in informational semantics by considering sequences of worlds that include all and only the worlds in an informational state. These sequences are what we call **paths**.

Paths resemble orderings that are familiar from standard truth-conditional semantics for conditionals—in particular, they are reminiscent of the total orderings that one can extract from Stalnaker's (1968) semantics. But paths are not meant to capture a notion of similarity and the context does not fix a 'path of evaluation'. Paths are theory-internal devices; actual utterances of conditionals are evaluated at information states. (Accordingly, intuitions about 'probabilities of conditionals' always refer to probabilities of conditionals at an information state, even though paths will be useful in calculating the latter.)

Formally, a path is a sequence of worlds without repetitions. Paths are related to information states: an information state fixes a set of paths. For a toy example, consider an information state i that involves three worlds:

$$i: \quad \{w_1, w_2, w_3\}$$

This information state fixes six paths:

$$\begin{gathered}\langle w_1, w_2, w_3\rangle\\ \langle w_1, w_3, w_2\rangle\\ \langle w_2, w_3, w_1\rangle\\ \langle w_2, w_1, w_3\rangle\\ \langle w_3, w_1, w_2\rangle\\ \langle w_3, w_2, w_1\rangle\end{gathered}$$

While truth at a path is the main compositional notion, the system also includes a notion of **support** of a sentence at an information state. A sentence is supported at an information state just in case it is true at all paths generated by that state.

Let us turn to stating the system more precisely. It's helpful to introduce two relations on paths. First, p' is a **subsequence** of p ($p' \leq p$) just in case every world in p'

[16]We have in mind in particular the theories that develop the initial idea in van Fraassen 1976; these include Bradley 2012, Kaufmann 2009, Bacon 2015. See also McGee 1989 and Jeffrey and Stalnaker 1994 for theories that involve related ideas.

[17]This idea builds on the semantics in Stalnaker 1968, 1981, 1984; see also Schein 2003, Schlenker 2004 for recent attempts in a similar vein.

is in p, and the worlds in p' appear in the same order as in p. For example, $\langle w_1, w_3, w_4\rangle$ is a subsequence of $\langle w_1, w_2, w_3, w_4\rangle$. Second, p' is a **permutation** of p ($p' * p$) just in case p' and p contain the exact same worlds, in a possibly different order. For example, $\langle w_2, w_1, w_4, w_3\rangle$ is a permutation of $\langle w_1, w_2, w_3, w_4\rangle$.

Finally, we introduce a notion of the **update** of a path p with a sentence A ($p + \mathsf{A}$), as the largest subsequence of p that makes A true throughout every permutation. For example, where A is true at w_2 and w_4 uniquely, $\langle w_1, w_2, w_3, w_4\rangle + \mathsf{A} = \langle w_2, w_4\rangle$.

Definition 1.

1. A **path** p is a sequence of worlds without replacement. A world w assigns a truth value $w(\mathsf{p})$ to every atomic sentence.

2. p_i is the i-th member of p.

3. An **interpretation function** $[\![\cdot]\!]$ assigns every sentence a set of paths.

4. $p' \leq p$ (p' is a **subsequence** of p) iff whenever w occurs earlier in p' than v, w occurs earlier in p than v.

5. $p' * p$ (p' is a **permutation** of p) iff p and p' consist of the same worlds, ordered potentially differently.

6. $p + \mathsf{A}$ (the **update** of p with A) is the largest member of the following set: $\{p' \leq p \mid \forall p'' \text{ if } p'' * p' \text{ then } p'' \in [\![\mathsf{A}]\!]\}$

In path semantics, an atom is true at a path just in case it is true at the first world in the path. The connectives are given the usual Boolean interpretation. The conditional is true at a path just in case the consequent is true at the result of updating the path with the antecedent. Finally, epistemic possibility and necessity modals existentially and universally quantify over permutations of paths.

Definition 2.

1. $[\![\mathsf{p}]\!]^p = 1$ iff $p_1(\mathsf{p}) = 1$

2. $[\![\neg\mathsf{A}]\!]^p = 1$ iff $[\![\mathsf{A}]\!]^p = 0$

3. $[\![\mathsf{A} \wedge \mathsf{B}]\!]^p = 1$ iff $[\![\mathsf{A}]\!]^p = 1$ and $[\![\mathsf{B}]\!]^p = 1$

4. $[\![\mathsf{A} \to \mathsf{B}]\!]^p = 1$ iff $[\![\mathsf{B}]\!]^{p+\mathsf{A}} = 1$

5. $[\![\Diamond\mathsf{A}]\!]^p = 1$ iff $\exists p' * p : [\![\mathsf{A}]\!]^{p'} = 1$

6. $[\![\Box\mathsf{A}]\!]^p = 1$ iff $\forall p' * p : [\![\mathsf{A}]\!]^{p'} = 1$

We define a notion of support at an information state. We also define two notions of consequence: one involves preservation of truth at a path, the other preservation of support.

Definition 3.

1. p supports A ($p \models \mathsf{A}$) iff $\forall p' * p : [\![\mathsf{A}]\!]^{p'} = 1$
2. $\mathsf{A}_1, \ldots, \mathsf{A}_n$ truth conditionally entails B ($\mathsf{A}_1, \ldots, \mathsf{A}_n \models_c \mathsf{B}$) iff $[\![\mathsf{B}]\!]^p = 1$ whenever $[\![\mathsf{A}_i]\!]^p = 1$ for every $i \in [1, n]$.
3. $\mathsf{A}_1, \ldots, \mathsf{A}_n$ informationally entails B ($\mathsf{A}_1, \ldots, \mathsf{A}_n \models_i \mathsf{B}$) iff $p \models \mathsf{B}$ whenever $p \models \mathsf{A}_i$ for every $i \in [1, n]$.

For a simple illustration, consider the path in (14).

(14) $\langle w_1, w_2, w_3 \rangle$

Let these worlds model the outcome of a fair die, where w_i is a world in which the die lands i. Then the following sentences are true at (14):

(15)
a. The die landed on 1.
b. The die landed on 1 or 3.
c. The die must not have landed on 5 or 6.
d. The die might have landed on 2.
e. If the die didn't land on 1, it landed on 2.

Let us also go through some examples of path updating. On our definition, the update of a path p with A is the largest subsequence of p such that all its permutations make A true. For reference, consider again a case where a die was tossed, and suppose that we are updating the following path:

$$p: \quad \langle w_1, w_2, w_3, w_4, w_5, w_6 \rangle$$

Here is how the update of this path works out in particular cases.

i. For any descriptive A, the update of any path p with A is just the subsequence of p consisting of A-worlds. For an example, let A be (16); p + (16) is in (17).

   (16) The die landed on an odd number.

   (17) p + (16): $\langle w_1, w_3, w_5 \rangle$

ii. Updating with $\Box\mathsf{A}$ has identical effects to updating with A (despite the differences in meaning between the two). Hence p + (18) is again the subsequence of p that includes all and only the odd worlds.

   (18) The die must have landed on an odd number.

   (17) p + (18): $\langle w_1, w_3, w_5 \rangle$

iii. Updating with $\mathsf{A} > \mathsf{C}$ has identical effects to updating with the material conditional $\mathsf{A} \supset \mathsf{C}$. So e.g. the update of p with (19) is in (17)

(19) If the die didn't land on two, it landed odd.

(17) p + (19): $\langle w_1, w_2, w_3, w_5 \rangle$

Before moving on, let us notice one feature of the semantics that will have consequences for a theory of probability. The notion of path update requires that, for every sentence A and for every path p, there is a unique largest subpath of p that makes A true. This condition is not met by some complex sentences. For example, sticking to the die scenario, consider (20):

(20) The die must have landed on two, or the die must have landed on an odd number.

If we update p with the first disjunct, we obtain $\langle w_2 \rangle$, and if we update p with the second disjunct, we obtain $\langle w_1, w_3, w_5 \rangle$. But there is no single largest subpath of p that makes true (20).

Santorio notices this problem, and builds a homogeneity requirement in the semantics of conditionals. Roughly: we admit that a sentence A can induce multiple largest updates of a path, and A > C is defined at p just in case C has the same truth value relative to all largest updates of p with A. We explain later on how this impacts the assignment of probabilities to conditionals.

This concludes our overview of path semantics. We now explain how probability and update can be defined for this semantics. We first give an overview in §5, and then introduce the details in later sections.

5 Overview of the proposal

In the next sections, we start from path semantics and develop a theory of probability and update that vindicates the **Update Thesis** and **Stalnaker's Thesis**$^-$. One important step in achieving this goal is to develop a new update operation, which we call 'Hyperconditionalization'. This section goes through an intuitive overview of how probability and update work in the theory.

Before proceeding, a point about notation. We will allow ourselves to be sloppy and use disjunctions of the form '$w_i \vee w_k \vee \ldots \vee w_n$' to stand for sentences true exactly at worlds $w_i, w_k, \ldots, w_n$. This mixes object language and metalanguage, but it allows us to be more concise.

5.1 Probability

Paths are the basic bearers of probability. To assign probability to paths, we start from an ordinary probability distribution over worlds and 'lift' it to paths. The lift operation is based on work by Justin Khoo (Khoo forthcoming) and is simple: the probability of a path is given by the product of the probabilities of each individual world, conditional on the previous worlds in the path being ruled out. Using a simple example, consider again the path $\langle w_1, w_2, w_3 \rangle$. Its probability is the product of the probability of w_1, the probability of w_2 conditional on w_1 not obtaining, and the probability of w_3 conditional on both w_1 and w_2 not obtaining. This way of assigning probability captures the idea of possibilities in the path as fallback options.

Paths are more fine-grained possibilities than worlds: each world corresponds to a set of paths, i.e. all the paths that have that world as their first member. Accordingly, while in standard models worlds are the basic points in an algebra, in our model worlds are treated as cells to be further divided into sub-possibilities which differ in their modal and conditional commitments.

Consider an information state with four worlds $\{w_1, w_2, w_3 w_4\}$. The resulting model will include four cells corresponding to each of w_1–w_4. In turn, each cell is divided into sub-cells corresponding to other fallback worlds in the path. For example, the cell corresponding to w_1 is divided into three sub-cells, depending on which world is the first fallback option (w_2, w_3, or w_4). Each of these sub-cells is further partitioned into two sub-cells, representing a choice of a second fallback world.

We can represent this in a diagram. For simplicity, we assume all of w_1–w_4 are equiprobable, so all cells have the same size. We use '$[w_1 \dots w_i w_k]$' to represent that w_k is being used as a fallback possibility with respect to $w_1 \dots w_k$ (more precisely, $[w_1 \dots w_k]$ represents the set of paths which begin with the sequence $\langle w_1, \dots, w_k \rangle$).

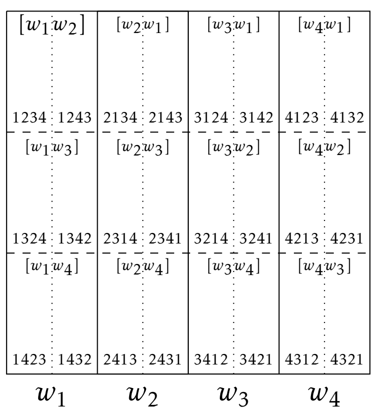

Figure 1: paths induced by the information state $\{w_1, w_2, w_3, w_4\}$

Suppose again that A is true at w_1, w_2 and B at w_2, w_3. The shaded area represents the portion of logical space where A → B is true.

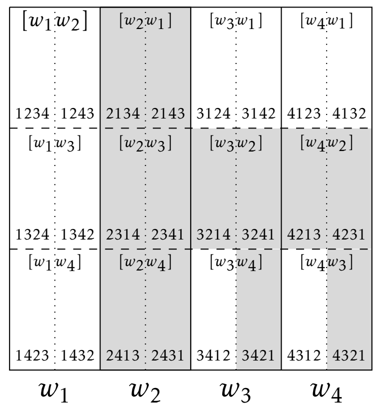

Figure 2: probability of $\mathsf{A} \to \mathsf{B}$, with $[\![\mathsf{A}]\!] = \{w_1, w_2\}$ and $[\![\mathsf{B}]\!] = \{w_2, w_3\}$.

This diagram helps illustrate how the account validates Stalnaker's Thesis$^-$. The area occupied by the conditional is half of the total area of the diagram. This accords with Stalnaker's Thesis, which predicts that $\mathsf{A} \to \mathsf{B}$ has probability $1/2 = \frac{Pr(\mathsf{A} \wedge \mathsf{B})}{Pr(\mathsf{A})} = \frac{1}{4}/\frac{1}{2}$.

5.2 Update

Conditionalization does not interact well with this picture of content. Suppose that, in our example, we conditionalize on $w_2 \vee w_3 \vee w_4$: hence we remove the entire w_1 cell, but leave the rest unaltered. Then w_1 still plays the role of fallback possibility:

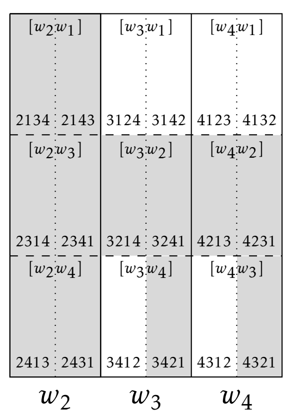

Figure 3: probability of $\mathsf{A} \to \mathsf{B}$ after conditionalizing on $w_2 \vee w_3 \vee w_4$.

This means that the conditional $\overline{w_2} \to w_1$ is assigned positive probability, despite the fact that w_1 has been ruled out. This will produce failures of the **Positive Preservation** condition discussed in §2.

To avoid this problem, we propose an update operation that we call **Hyperconditionalization**. Hyperconditionalization works as follows. We first update each path individually by removing worlds that are incompatible with the new information. For example, suppose again that we update on $w_2 \vee w_3 \vee w_4$. The paths on the left are updated to the paths on the right.

$$\begin{array}{lcl} \langle w_3, w_2, w_1, w_4 \rangle & \Rightarrow & \langle w_3, w_2, w_4 \rangle \\ \langle w_2, w_1, w_4, w_3 \rangle & \Rightarrow & \langle w_2, w_4, w_3 \rangle \\ \langle w_1, w_2, w_3, w_1 \rangle & \Rightarrow & \langle w_2, w_3, w_4 \rangle \\ \dots & \Rightarrow & \dots \end{array}$$

Then each path simply transfers its credence to the path that results from updating it. Given the way that paths are constructed, this transfer is many-to-one: several paths from the previous epistemic state transfer their credence to the same path in the updated epistemic state.

To illustrate this process, consider Figure 4. The arrows represent the transfer of probability from any given path to its resulting path in the result of hyperconditionalizing on $w_2 \vee w_3 \vee w_4$. To avoid clutter, we only represent w_1 and w_2; w_3 and w_4 work pretty much like w_2.

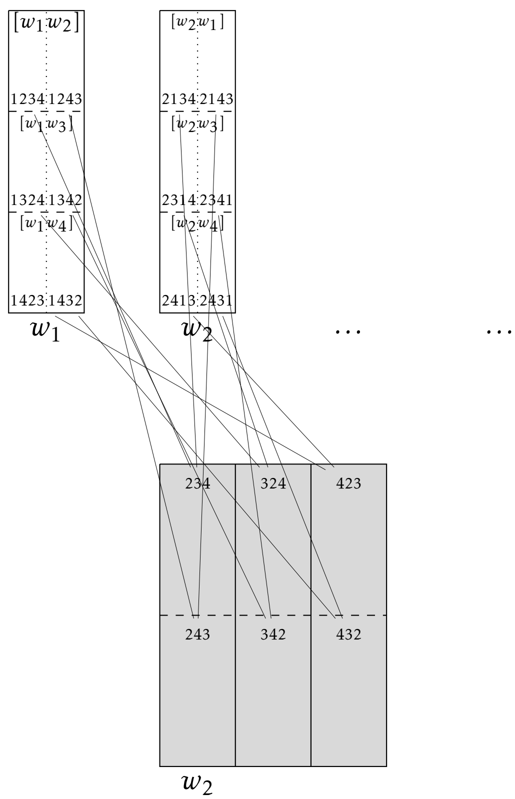

Figure 4: probability of $\mathsf{A} \to \mathsf{B}$ after update on $\overline{w_1}$.

The crucial feature of this procedure is that it makes the update operation and the lift operation commutative. Suppose we start from an ordinary credence distribution Pr over worlds, and a proposition p that is the new information. Now consider two different ways of proceeding: (i) we lift Pr to an epistemic state involving a credence distribution over paths, and hyperconditionalize the latter on p; (ii) we conditionalize

Pr on p, and then lift the result to an epistemic state involving a credal distribution over paths. Procedure (i) and (ii) yield the same results. This means that Hyperconditionalization is in some sense conservative with respect to classical probability.

This ends our overview; we now turn to spelling out the theory in detail.

6 Probability

The starting idea for our account is that each agent is endowed with two credence functions. The first is an ordinary credence function, which assigns probabilistic credences to standard propositions. The second credence function assigns credal values to all contents, including contents of modal and conditional claims. Crucially, this second credence function can be derived from the first via a 'lifting' operation that we describe below. So there is a sense in which the second credence function is grounded in the subject's credences over standard contents.

Even though we can ultimately rely on standard credences to fix credences in all contents, both credence functions model genuine attitudes. Lifted credences are no less 'real' than basic credences. In particular, lifted credences are crucial for modeling attitudes towards conditionals. These attitudes arguably play a key role in a number of enterprises: for example, decision theory and explaining subject's behavior.[18]

In this section, we introduce the two credence functions and the lifting procedure. After that, we show that the lifting operation guarantees that conditionals are assigned probabilities that conform to Stalnaker's Thesis. Our proof is an adaptation of a proof by Justin Khoo (in Khoo forthcoming), which in turn modifies proofs about probabilities of sequences that are based on work by Bas Van Fraassen (van Fraassen 1976, Bradley 2012, Kaufmann 2009, Bacon 2015).

6.1 Epistemic spaces and proto epistemic spaces

We assume that every agent is endowed with two different credence measures: one over possible worlds, and one over possible paths. The former makes up what we call the agent's "proto epistemic space (variable: U)". The latter makes up we call an "epistemic space" (variable: E). Both proto epistemic spaces and epistemic spaces are pairs of a set of possibilities, and a probability measure defined over the power set (the set of all subsets) of those possibilities. In both cases, for the sake of simplicity, we will assume that the set of possibilities is finite and that the probability distribution is regular, assigning a positive probability to each point.[19]

[18]Some theorists suggest introducing the notion of a *suppositional attitude* and using it in decision theory and related enterprises (see e.g. Williams 2012). For example, on Williams' view, Sarah has a suppositional credence in the die landing 6, on the supposition that it lands even. Crucially, though, on an account like Williams' suppositional credences are supposed to march in step with credences in the corresponding conditionals.

[19]We depart slightly from a more standard presentation in which a space is a triple of a set of possibilities, an algebra defined on that set, and a probability measure defined on the algebra. We grant ourselves

Definition 4. Proto-epistemic space.

A proto epistemic space is a pair $U = \langle W, Pr \rangle$, where:

1. W is a set of possible worlds.

2. Pr assigns a positive probability to every member of the power set of W.

The agent's proto epistemic space is then used to construct her credence function over paths, her epistemic space (E). Since the meaning of each sentence in our language is a set of paths, this new credence function determines an agent's credence in any particular claim.

Definition 5. Epistemic space.

An epistemic space is a pair $E = \langle P, C \rangle$, where:

1. P is a set of paths.

2. C assigns a probability to every member of the power set of P.

Drawing on Khoo (in Khoo forthcoming), we introduce a procedure for deriving an epistemic space from a proto epistemic space. The idea is that every probability measure over worlds induces a unique probability measure over paths, according to a precise definition. To state this definition, it is useful to introduce a bit of notation for certain special sets of paths that agree on a fixed initial segment of worlds. To that end, we let $p[w_1, \ldots, w_n] = \{p \mid p_1 = w_1, \ldots, p_n = w_n\}$. Now we introduce our preferred operation $\uparrow$ for deriving a probability measure over paths from a measure over worlds.

Definition 6. Lift to epistemic spaces.

The epistemic space $E = \uparrow U$ is lifted from proto-epistemic space $U = \langle W, Pr \rangle$ iff $E = \langle P, C \rangle$ where:

1. P is the set of all paths of worlds in W.

2. C assigns a probability to every member of the power set of P, subject to the constraint:

   (a) $C(p[w]) = Pr(w)$

   (b) $C(p[w_1, \ldots, w_n]) = C(p[w_1, \ldots, w_{n-1}]) \times \frac{Pr(w_n)}{Pr(W - \{w_1, \ldots, w_{n-1}\})}$

With our lift operation defined, we can then define the class of epistemic spaces over paths that can be derived from some underlying measure over worlds.

Definition 7. Well-behaved epistemic spaces.

this simplification because we assume in each case that the relevant algebra is simply the power set of the possibilities under discussion.

E is a well-behaved epistemic space just in case there exists some proto epistemic space U where $E = \uparrow U$.

The notion of well-behavedness plays a crucial role in our account, and in particular in our proof of ST$^-$. In this section, we show that every well-behaved credence distribution vindicates ST$^-$. In the next section, we show that our proposed update operation, differently from standard Conditionalization, maps well-behaved spaces to well-behaved spaces. So well-behavedness figures prominently in the proofs of the two main formal results (i.e. Theorems 1 and 3 below).
To illustrate the idea of path probability, let's return to our working example. Now imagine that our three possible worlds are equipped with a probability measure, and consider how that would affect the corresponding probability of the resulting paths. From the construction above, we have

$$C(\langle w_1, w_2, w_3\rangle) = C(p[w_1, w_2]) \times \frac{Pr(w_3)}{Pr(w_3)} = C(p[w_1, w_2]) =$$

$$= C(p[w_1]) \times \frac{Pr(w_2)}{Pr(\{w_2, w_3\})} = Pr(w_1) \times \frac{Pr(w_2)}{Pr(\{w_2, w_3\})} = \frac{Pr(w_1) \times Pr(w_2)}{Pr(\{w_2, w_3\})} = \frac{Pr(w_1) \times Pr(w_2)}{Pr(w_2) + Pr(w_3)}$$

Generalizing, we have that $C(\langle w_a, w_b, w_c\rangle) = \frac{Pr(w_a) \times Pr(w_b)}{Pr(w_b) + Pr(w_c)}$. The probability of path $\langle w_a, w_b, w_c\rangle$ is found by weighting w_a's probability by the amount of probability w_b is assigned conditional on w_a not obtaining. This induces the following stochastic truth table, filling in a particular choice of probability measure for worlds:

world	Pr
w_1	3/6
w_2	2/6
w_3	1/6

path	C
w_1, w_2, w_3	2/6
w_1, w_3, w_2	1/6
w_2, w_1, w_3	3/12
w_2, w_3, w_1	1/12
w_3, w_1, w_2	3/30
w_3, w_2, w_1	2/30

Notice that in this example, the ur-probability of w_1 is $\frac{3}{6}$, and this is also the probability of the set of all paths that begin with w_1. This points to a more general structural feature. When E is constructed from U, E and U assign the same probability to any descriptive claim. After all, descriptive claims are true at a path just when they are true at the first world in the path. So the meaning of any descriptive claim is the set of paths where that claim is true at the first world in it. But the construction procedure above for deriving E from U guarantees that the probability of any set of paths that agree on the initial world is simply the probability of that world in U.

6.2 Vindicating Stalnaker's Thesis$^-$

Now that we've seen an example of path probability, we turn to its signature property: the restricted version of Stalnaker's Thesis, i.e. **ST**$^-$, is valid.

Theorem 1. For all C, for all descriptive A and B, and for any C that is the credence function of a well-behaved epistemic state E:

$$C(\mathsf{A} \to \mathsf{B}) = C(\mathsf{B} \mid \mathsf{A})$$

We confine the full proof to an appendix. But we can illustrate the key ideas through our running example.

world	Pr
w_1	3/6
w_2	2/6
w_3	1/6

path	C
w_1, w_2, w_3	2/6
w_1, w_3, w_2	1/6
w_2, w_1, w_3	3/12
w_2, w_3, w_1	1/12
w_3, w_1, w_2	3/30
w_3, w_2, w_1	2/30

Now consider the probability of the following conditional:

(21) $(w_1 \vee w_2) \to w_1$
If the die didn't land 3, it landed 1.

This conditional states that if an agent is in one of w_1 or w_2, they are in w_1. Stalnaker's Thesis says that the probability of this conditional is simply the probability of being in w_1, conditional on being in w_1 or w_2. Now (21) is true at any path where w_1 precedes w_2. So it is true at $\langle w_1, w_2, w_3\rangle$, $\langle w_1, w_3, w_2\rangle$, and $\langle w_3, w_1, w_2\rangle$. The probability of this set of paths is just their sum: $\frac{2}{6}+\frac{1}{6}+\frac{3}{30} = \frac{3}{5}$. Now consider the conditional probability of w_1 given $w_1 \vee w_2$. w_1 is true at all paths whose first member is w_1: this is $\langle w_1, w_2, w_3\rangle$ and $\langle w_1, w_3, w_2\rangle$. By contrast, $w_1 \vee w_2$ is true at the first four paths, where the first world in the path is either w_1 or w_2. So the relevant conditional probability is $\frac{P(p[w_1])}{P(p[w_1] \cup p[w_2])} = \frac{P(\{\langle w_1,w_2,w_3\rangle, \langle w_1,w_3,w_2\rangle\})}{P(\{\langle w_1,w_2,w_3\rangle, \langle w_1,w_3,w_2\rangle, \langle w_2,w_1,w_3\rangle, \langle w_2,w_3,w_1\rangle\})} = \frac{2/6+1/6}{2/6+1/6+3/12+1/12} = \frac{1/2}{5/6} = \frac{3}{5}$. Here we have a valid instance of Stalnaker's Thesis.

In passing, let us also notice that the foregoing also shows that our theory vindicates **Positive Preservation**. As we noted, **Positive Preservation** is entailed by ST, hence by vindicating the latter we also vindicate the former.

We've now developed a theory of probability that vindicates Stalnaker's Thesis. To avoid triviality, we introduce a new update procedure that replaces conditionalization. Before we turn to that task, however, we pause to note that our theory also makes reasonable predictions about the probability of epistemic modal claims. [CHECK]

6.3 Vindicating Must Preservation, Might Contradiction, and Positive Preservation

In path semantics, epistemic modals quantify over permutations of paths. $\Box \mathsf{A}$ is true at p just in case A is true at every permutation of p. $\Diamond \mathsf{A}$ is true at p just in case A is true at some permutation of p. These meanings correspond to the standard ones in an information sensitive semantics: $\Box \mathsf{A}$ and $\Diamond \mathsf{A}$ are true at an information state i iff, respectively, all or some world in i validates A.

This treatment of epistemic modals induces a transparent theory of their probability. The probability of a modal claim is always 1 or 0, depending on the probabilistic status of that claim's prejacent.

Theorem 2. For any A and for any C that is the credence function of a well-behaved epistemic state E:

1. $C(\Box \mathsf{A}) = \begin{cases} 1 & \text{if } C(\mathsf{A}) = 1 \\ 0 & \text{otherwise} \end{cases}$

2. $C(\Diamond \mathsf{A}) = \begin{cases} 1 & \text{if } C(\mathsf{A}) > 0 \\ 0 & \text{otherwise} \end{cases}$

An immediate consequence of Theorem 2 is that we vindicate both of the principles about epistemic modals that gave rise to triviality, i.e. **Must Preservation** and **Might Contradiction**. Stated in terms of lifted credences, these principles say:

> **Must Preservation (MP).** Where C is a rational credence function: If $C(\mathsf{A}) = 1$, then $C(\Box \mathsf{A}) = 1$.
>
> **Might Contradiction (MC).** Where Pr is a rational probability function: If $C(\bar{\mathsf{A}}) = 1$, then $C(\Diamond \mathsf{A}) = 0$.

The inference from A to $\Box \mathsf{A}$ preserves certainty: whenever an agent is certain of A, their credence in $\Box \mathsf{A}$ is 1. But the inference does not preserve probability. If the agent's credence in A is merely $\frac{2}{3}$, the transparency thesis implies that their credence in $\Box \mathsf{A}$ is 0. Similarly, an agent's being certain of $\bar{\mathsf{A}}$ guarantees that they assign zero credence to $\Diamond \mathsf{A}$. But a rational agent who has credence less than 1 in $\bar{\mathsf{A}}$ will be certain that $\Diamond \mathsf{A}$.

We confine the proof of Theorem 2 to an appendix. To get a feeling for why it holds, consider our running example.

world	Pr
w_1	3/6
w_2	2/6
w_3	1/6

path	C
w_1, w_2, w_3	2/6
w_1, w_3, w_2	1/6
w_2, w_1, w_3	3/12
w_2, w_3, w_1	1/12
w_3, w_1, w_2	3/30
w_3, w_2, w_1	2/30

Against the background of this scenario, consider first:

(22) a. The die landed on 1, 2, or 3.
b. The die must have landed on 1, 2, or 3.

(23) a. The die didn't land on 4.
b. The die might have landed on 4.

The statements in (22) witness **Must Preservation**, since they both have probability 1. The statement in (23) witness **Might Contradiction**, since (23-a) has probability 1 and (23-b) has probability 0.

At the same time, we can see that **Must Preservation** and **Might Contradiction** don't, in general, preserve probability (or lack thereof). Consider

(24) a. The die landed 1.
b. The die might have landed 1.
c. The die must have landed 1.

C(The die landed 1) is the probability of the set of paths at which the die lands 1. This set is $\{\langle w_1, w_2, w_3\rangle, \langle w_1, w_3, w_2\rangle\}$. So we have:

$$C(\text{The die landed 1}) = 2/6 + 1/6 = 1/2$$

Differently from (24-a), (24-b) and (24-c) are true at all paths or at none; hence they are always assigned extreme probabilities. In this particular case, we have:

$$C(\text{The die might have landed 1}) = 1$$
$$C(\text{The die must have landed 1}) = 0$$

Hence, when A has intermediate credence, $\Diamond\mathsf{A}$ has credence 1, and $\Box\mathsf{A}$ has credence 0.

Assigning always only extreme probabilities to *might* A and *must* A is controversial (see e.g. Moss 2015, Charlow Forthcoming). Let us emphasize that this part of the framework can be disentangled from the rest: a different semantics for modals would yield different results. Given constraints of space, we won't consider alternative options here.[20]

7 Update

This section develops our rule for rational update. First, we review why conditionalization is incompatible both with the way in which we construct epistemic spaces, and with Stalnaker's Thesis. Then we develop our alternative.

[20]For example, we could enrich the theory above with the generalized accessibility relations in Goldstein 2018.

7.1 Why simple conditionalization won't work

Suppose an agent's current epistemic space is $\langle P, C\rangle$. The first, obvious proposal for modeling update is that the agent's updated credence function after learning A is $C(\cdot \mid \mathsf{A})$, which assigns any claim B the ratio of $C(\mathsf{A} \wedge \mathsf{B})$ to $C(\mathsf{A})$.

Definition 8. Conditionalization on epistemic spaces.
Where epistemic space $E = \langle P, C\rangle$, $E +_{\mathsf{C}} \mathsf{A} = \langle P, C^{+}\rangle$, where $C^{+}(\mathsf{B}) = \frac{C(\mathsf{B} \wedge \mathsf{A})}{C(\mathsf{A})}$.

The problem is that this operation does not always return a well-behaved epistemic space. I.e., the resulting model will not be lifted from a probability distribution on worlds. To see this, return to our running example:

world	Pr
w_1	3/6
w_2	2/6
w_3	1/6

path	C
w_1, w_2, w_3	2/6
w_1, w_3, w_2	1/6
w_2, w_1, w_3	3/12
w_2, w_3, w_1	1/12
w_3, w_1, w_2	3/30
w_3, w_2, w_1	2/30

Suppose we conditionalize C on the information that w_1 doesn't obtain, by removing all paths starting with w_1. We reach the following probability measure:

path	C^{+}
w_1, w_2, w_3	0
w_1, w_3, w_2	0
w_2, w_1, w_3	1/2
w_2, w_3, w_1	1/6
w_3, w_1, w_2	1/5
w_3, w_2, w_1	2/15

The probability measure in C^{+} has two undesirable properties. First, it cannot result from lifting a proto-epistemic space $\langle W, Pr\rangle$. Since the probability of path $\langle w_1, w_2, w_3\rangle$ is 0, given the way path probability is defined, Pr must assign probability zero to one of w_1, w_2, or w_3. But this would imply that the probability of all the other paths containing these worlds would also be zero. Second, C^{+} does not vindicate Stalnaker's Thesis. Notice that $C^{+}(w_2 \vee w_3 \mid w_1 \vee w_3) = 1$ (since $w_2 \vee w_3$ has a probability of 1). But the conditional $(w_1 \vee w_3) \rightarrow (w_2 \vee w_3)$ is false at a path that has positive probability, namely $\langle w_2, w_1, w_3\rangle$, hence the conditional gets probability lower than 1.

7.2 Hyperconditionalization

We say that while $\langle P, C\rangle$ is a rational epistemic space, $\langle P, C\rangle +_{\mathsf{C}} \mathsf{A}$ is not. The class of rational epistemic spaces is not closed under conditionalization. Instead, it is closed under a new update rule, which we call 'Hyperconditionalization' and represent as '$+_{\mathsf{HC}}$'. Hyperconditionalization is different extensionally from Conditionalization, but it is conceptually similar. The basic idea is that possibilities incompatible with the new information should be ruled out from appearing at any position in paths. Accordingly, we cannot merely rule out paths, but we also have to remove some of their members.[21] Crucially, this rule will take well-behaved epistemic spaces to well-behaved epistemic spaces.

We start by giving a concrete illustration of Hyperconditionalization, and then state the rule formally.

To start, recall the notion of path update:

> $p + \mathsf{A}$ (the **update** of p with A) is the largest member of the set:
>
> $$\{p' \leq p \mid \forall p'' \text{ if } p'' * p' \text{ then } p'' \in [\![\mathsf{A}]\!]\}$$

I.e., the update of a path p with A is the largest subsequence of p such that all of its permutations make A true. We gave some examples of path update in §4. Hyperconditionalizing an epistemic state E on A proceeds in two simple steps.

i. First, for every path p in E, we determine its update $p + \mathsf{A}$.

ii. Second, for every path p in E, we transfer its credence to $p + \mathsf{A}$.

Let us illustrate this with our working example. Suppose that the agent starts from the following credence distribution in worlds and resulting credence function over paths.

world	Pr
w_1	3/6
w_2	2/6
w_3	1/6

path	C
w_1, w_2, w_3	2/6
w_1, w_3, w_2	1/6
w_2, w_1, w_3	3/12
w_2, w_3, w_1	1/12
w_3, w_1, w_2	3/30
w_3, w_2, w_1	2/30

Now suppose that the agent learns (25).

(25) The die landed on 2 or 3.

As a first step, she updates each path with the information in (25).

[21] Hyperconditionalization is similar to the update operation of 'deep conditioning' in the Van Fraassen-style framework used by Kaufmann 2015.

$$\begin{array}{rcl}
\langle w_1, w_2, w_3\rangle & \Rightarrow & \langle w_2, w_3\rangle \\
\langle w_1, w_3, w_2\rangle & \Rightarrow & \langle w_3, w_2\rangle \\
\langle w_2, w_1, w_3\rangle & \Rightarrow & \langle w_2, w_3\rangle \\
\langle w_2, w_3, w_1\rangle & \Rightarrow & \langle w_2, w_3\rangle \\
\langle w_3, w_1, w_2\rangle & \Rightarrow & \langle w_3, w_2\rangle \\
\langle w_3, w_2, w_1\rangle & \Rightarrow & \langle w_3, w_2\rangle
\end{array}$$

Then she transfers the credence in the paths in her old epistemic state to the updated paths. This mapping is many-to-one, so the credence in the new paths ends up being a sum of credences in the old paths. This is the result we get:

path	C^*
w_2, w_3	$2/6 + 3/12 + 1/12 = 2/3$
w_3, w_2	$1/6 + 3/30 + 2/30 = 1/3$

We now give a formal statement of the update operation.

Definition 9. Hyperconditionalization.
Where epistemic space $E = \langle P, C\rangle$, $E +_{\mathsf{HC}} \mathsf{A} = E_{\mathsf{A}} = \langle P_{\mathsf{A}}, C_{\mathsf{A}}\rangle$, where:

(i) $P_{\mathsf{A}} = \{p + \mathsf{A} \mid p \in P\}$

(ii) $C_{\mathsf{A}}(p) = \sum C(p') : p = p' + \mathsf{A}$

(iii) $C_{\mathsf{A}}(\mathsf{B}) = \sum_{p \in \mathsf{B}} C_{\mathsf{A}}(p)$

This concludes our statement of Hyperconditionalization. We now turn to showing that Hyperconditionalization validates the **Update Thesis** and **Robust Stalnaker**.

7.3 Proving the Update Thesis

Having defined both a semantics for conditionals and a full theory of probability and update, we are in a position to prove the **Update Thesis**. Stated in terms of lifted credences, the Thesis says the following:

Update Thesis. For all rational C and all A, B: $C(\mathsf{A} \to \mathsf{B}) = C_{\mathsf{A}}(\mathsf{B})$

The proof works simply by observing that, given our semantics and our accounts of probability and update, probabilities of conditionals and updated probabilities are equal by construction.

First of all, notice that a subject S's credence in $\mathsf{A} \to \mathsf{C}$ equals the sum of S's credences in the paths at which $\mathsf{A} \to \mathsf{C}$ is true. The latter sum, in turn, equals the sum of S's credences in the paths that, updated with A, make C true.

$$C(\mathsf{A} \to \mathsf{C}) \quad = \quad \sum_{p \,\vDash\, \mathsf{A} \to \mathsf{C}} C(p) \quad = \quad \sum_{p + \mathsf{A} \,\vDash\, \mathsf{C}} C(p)$$

Consider now S's credence in C, updated in A. Notice first that (similarly to what happened above) this credence equals to the sum of S's credences, updated on A in paths that make C true.

$$C_{\mathsf{A}}(\mathsf{C}) \quad = \quad \sum_{p \,\vDash\, \mathsf{C}} C_{\mathsf{A}}(p)$$

But now, via clause (ii) of Definition 9, we know that S's credence, updated on A in a path p equals the sum of S's previous credences in paths p' such that p', updated on p, is identical to p.

$$C_{\mathsf{A}}(p) \quad = \quad \sum_{p':\ p=p'+\mathsf{A}} C_{\mathsf{A}}(p')$$

Putting these two together, we get that S's credences in C, updated on A, equal the sum of their non-updated credences in paths that, updated with A, make C true.

$$C_{\mathsf{A}}(\mathsf{C}) \quad = \quad \sum_{p':\ \exists p:\ p=p'+\mathsf{A} \wedge p \,\vDash\, \mathsf{C}} C_{\mathsf{A}}(p') \quad = \quad \sum_{p'+\mathsf{A} \,\vDash\, \mathsf{C}} C_{\mathsf{C}}(p')$$

Putting all together, we get that S's credence in $C(\mathsf{A} \to \mathsf{C})$ and their updated credence $C_{\mathsf{A}}(\mathsf{C})$ both equal the same sum of credences in paths.

$$C(\mathsf{A} \to \mathsf{C}) \quad = \quad \sum_{p+\mathsf{A} \,\vDash\, \mathsf{C}} C(p) \quad = \quad C_{\mathsf{A}}(\mathsf{C})$$

The **Update Thesis** is one of the three main results of this paper, so it may seem surprising that the proof is so unremarkable. But the reason why the proof is unremarkable is that the definitions of update used in the semantics and in the model of credence are close variants of each other. This is just a formal implementation the insight at the basis of the suppositional view: there is a close link between the update performed by conditional *if*-clauses and the update of epistemic states.

7.4 Proving Robust Stalnaker

Our third *desideratum* for a theory of conditionals and probability is what we called **Robust Stalnaker**. Stated in terms of lifted credences, **RS** says:

Robust Stalnaker. For all rational C, all descriptive A, B, and all C:

$$C_{\mathsf{C}}(\mathsf{A} \to \mathsf{B}) = C_{\mathsf{C}}(\mathsf{B} \mid \mathsf{A})$$

RS says that by updating a credence function on any sentence C, we reach a new credence function on which the credence assigned to a conditional $\mathsf{A} \to \mathsf{C}$ equals the conditional credence $C_{\mathsf{C}}(\mathsf{B} \mid \mathsf{A})$. Robust Stalnaker amounts to saying that **Stalnaker's Thesis**$^{-}$ is robust under update.

To prove **RS**, we take a detour through another property of the system. We first notice that Hyperconditionalization is equivalent to a different way of updating epistemic spaces. Then, we point out that this guarantees that **Stalnaker's Thesis**$^{-}$ will hold at the updated distribution.

Suppose that, rather than hyperconditionalizing on their lifted credence function, the subject updates as follows. First, they conditionalize their proto-epistemic state on the relevant proposition.[22] Then, they lift the updated proto-epistemic state to an epistemic state. As it turns out, the two update procedures yield identical results. This is captured by our Theorem 3.

Theorem 3. Let $E = \langle P, C \rangle$ be an epistemic state such that $E = \uparrow U$, with $U = \langle W, Pr \rangle$. Let $E^{\mathsf{A}} = \langle P^{\mathsf{A}}, C^{\mathsf{A}} \rangle = \uparrow U_{\mathsf{A}}$, where U_{A} is the proto-epistemic space $\langle W \cap [\![\mathsf{A}]\!], Pr_A \rangle$ we reach by conditionalizing U on A. Then:

$$C^{\mathsf{A}}(p) \quad = \quad \sum C(p') : p = p' + \mathsf{A} \quad = \quad C_{\mathsf{A}}(p)$$

The proof of Theorem 3 is confined to the Appendix.

Theorem 3 is philosophically significant. It shows that we can view an agent's credences over paths as evolving in two different ways. First, we can think exclusively about her credences in paths, and represent her as hyperconditionalizing. Alternatively, we can represent her as fundamentally updating her credences in worlds, and then deriving a new path by our lift operation.

This shows that we can think of our new update rule as emerging naturally out of more basic ingredients. Fundamentally, an agent is endowed with a credence function over worlds. This credence function changes over time by the familiar process of ordinary conditionalization. But at any time, an agent's doxastic perspective extends beyond her credence in worlds. Her credence in worlds determines her broader outlook on modal matters via our construction procedure. And her credences in the latter count: as we mentioned, credences in conditionals will presumably play a role in decision theory and the explanation of behavior. Hyperconditionalization shows how the agents' credences should update when we take the modal perspective.

Theorem 3 also helps us prove two further results. As anticipated, the first is **Robust Stalnaker**. To see this, notice first that the following is a consequence of Theorem 3:

Theorem 4. Suppose E is well-behaved. Then, for all A, E_{A} is well-behaved.

An epistemic state is well-behaved iff it is the result of lifting a proto-epistemic space. Suppose that E is well-behaved, and hence there is a proto-epistemic state U that it is lifted from. Theorem 3 states that E_{A} is equivalent to the result of conditionalizing U on A, and then lifting the resulting proto-epistemic state. This guarantees that there is a proto-epistemic state that, when lifting, yields E_{A}. Hence E_{A} is well-behaved.

Now, via **Stalnaker's Thesis**$^{-}$, we know that, in every well-behaved epistemic state, if A and B are descriptive, $C(\mathsf{A} \rightarrow \mathsf{B}) = C(\mathsf{A} \mid \mathsf{B})$. Hence Theorem 4 and **Stal-**

[22] Some work will need to go into defining what propositions are used for update in path semantics; see Santorio 2019b for details. The basic idea is that, in analogy with path update, when updating i on A, the proposition we conditionalize on is the largest subset of i, $i' \subseteq i$ such that A is true at all paths generated from i'.

naker's Thesis$^-$ together guarantee that, whenever we update a lifted credence distribution on any sentence, we reach a credence distribution that validates **Stalnaker's Thesis$^-$**. This is **Robust Stalnaker**.

Theorem 3 also explains why **Hyperconditionalization** delivers the same results as **Conditionalization** for the descriptive fragment of the language. We saw earlier that whenever E is constructed from U, E and U agree on the probability of any descriptive claim. But Theorem 3 shows that hyperconditionalizing delivers the same results as simply conditionalizing U and constructing a new epistemic space. For this reason, when E is hyperconditionalized on A we know that the result will assign any descriptive claim B the conditional probability of B given A in U.

8 Further issues

8.1 Hyperconditionalization as Imaging

Before moving on, one quick remark. Hyperconditionalization is an instance of a well understood alternative to conditionalization: imaging. When an agent updates with A via imaging, she shifts probability from one set of points to another with the help of a selection function.[23] The probability of any point x is shifted to $f(x, \mathsf{A})$, the point that is in some sense the most similar to x where A is true. Our update rule is a type of imaging. Our update rule functions by shifting probability from a set of paths to another set of 'selected paths'. In particular, for us $f(p, \mathsf{A}) = p + \mathsf{A}$. Crucially, however, our rule differs from the kind of imaging discussed in extant literature, because the selection function implicit in our update procedure operates on paths rather than worlds. Empirically, this means that our update rule validates Stalnaker's Thesis, which connects the probability of the conditional to the corresponding conditional probability, rather than simply to the probability of the consequent after imaging on the antecedent. By contrast, consider an ordinary possible worlds semantics for conditionals in terms of selection functions. If probabilities are assigned to sets of worlds and an agent updates her credences by imaging, her credence in conditionals will not satisfy Stalnaker's Thesis. In this respect, our theory departs significantly from previous applications of imaging.

8.2 Undefined updates

The definition of hyperconditionalization exploits the notion of path update. There is a wrinkle though. Recall from §4 that the notion of the update of a path with a sentence A is not well defined for every sentence. One relevant example is the following:

(20) The die must have landed on two, or the die must have landed on an odd number.

[23] See for example Lewis 1976, Gardenfors 1982.

Path semantics solves this problem by building a homogeneity requirement in the semantics of conditionals. $A > C$ is defined just in case C has the same truth value on all the largest updates of p with A. This solves the problem for defining truth (at the cost of moving to a trivalent system), but not the problem of defining probability.

We see a few solution to address this problem.[24] The one that we adopt is simply to accept that probabilities for some conditionals with complex antecedents are undefined. We choose this option because it strikes us as very plausible empirically. For an example, consider:

(26) If the die must have landed on two or it must have landed on an odd number, then it landed on 1, 2, or 3.

What is the probability of (26)? The answer is unclear. It is very hard to form any kind of precise judgment. So we find it plausible to say that (26) does not have a precise probability, and to propose a theory that predicts that.[25]

8.3 Updating on modal information

Our theory also delivers interesting predictions about update with modal claims. Recall two features of path update from §4:

i. Updating a path with $\Box A$ has identical effects to updating it with A

ii. Updating a path with $A \to C$ has identical effects to updating it with the material conditional $A \supset C$.

Given these features, the credal update on an epistemic states that is generated by $\Box A$ and $A > C$ is fully fixed: we just look at the results of hyperconditionalizing on, respectively, A and $A \supset C$.

For the case of $\Diamond A$, things are slightly different. This claim functions as a test, in the sense of Veltman 1996. Updating a path on $\Diamond A$ either maintains the path unaltered, or removes all the worlds from the path, effectively taking the path to an 'absurd' state. Things will work analogously for credence. Either the subject's credal distribution will remain the same, or the subejct will end up with an absurd credence distribution and will have to perform some belief revision.

[24] One option would be to impose a symmetric requirement for probability: the probability of $A > C$ is defined at p iff, for every maximal update A' of p relative to A, $p[A']$ receives the same probability. This might work, but it would still have the effect that, in the great majority of cases, the probability of $A > C$ is undefined when A does not induce a unique maximal update. Another viable option would involve using as the update of p relative to A the union of all the maximal updates as we have defined them. As an anonymous referee points out, this comes with some drawbacks for the logic: in particular, it invalidates the idempotence principle $A > A$.

[25] A variant of this proposal: conditionals with antecedents that don't induce a unique multiple update have interval-valued probabilities. Before launching into developing this proposal formally, it seems advisable to gather some further empirical evidence for the claim that conditionals have interval-valued probabilities.

This is a good moment to point out that the overall view of modal content and update is not committed to the specific views that $\Box A$ and $\Diamond A$ are always assigned extreme probabilities. This is just the simplest option to take, and what is suggested by adopting the most basic version of informational semantics for modals. But there are arguments to think that *might*- and *must*-claims have a more sophisticated semantics; see e.g. Willer 2013, Moss 2015, Goldstein 2018. These views are in principle compatible with out apparatus, though for reasons of space we have to leave discussion of them to future work.

8.4 Consequences: Assertability and logical consequence

Our theory has a consequence that is, at first sight, surprising: sentences that are informationally incompatible are probabilistically compatible. To see this, go back to the die example; a six-sided die has been tossed, and we have no information about the outcome. Consider:

(27) a. The die didn't land on 6.
b. The die might have landed on 6.

Holding fixed the toy model where we have six worlds, each with equal credence, (27-a) and (27-b) are assigned, respectively, probability 5/6 and 1. What is more striking, their conjunction (in (28)), which is a so-called epistemic contradiction and informationally inconsistent, is also assigned probability 5/6.

(28) The die didn't land on 6 and it might have landed on 6.

This might give rise to two potential worries.[26] The first has to do with assertability, and the second with consequence.

The first worry is that this result is clearly problematic if we hold, following a suggestion going back to Adams 1975, that credence tracks (at least, roughly) the degree of assertability of a proposition. Adams' idea has two components. The first is that the assertability of a sentence is not an all-or-nothing issue, but rather comes in degrees. The second is that degrees of assertability match degrees of truth. This idea has a fruitful application just to conditionals: the degree of assertability of a conditional $A \rightarrow B$ equals the conditional probability of B, given A.

Now, preserving Adams' idea would have disastrous consequences for us, in light of the facts discussed above. We would predict that epistemic contradictions like (28) are highly assertable. But of course, whether we regard (28) as infelicitous for semantic or pragmatic reasons, (28) is obviously infelicitous. Our first response is that degree of credence simply doesn't track degree of assertability. We are inclined to accept a knowledge norm on assertion (Williamson 2000). But an agent's credence in p can be quite high without her knowing p. For example, an agent can have a quite high credence that she will lose the lottery, and yet this claim is not assertable to any

[26] Thanks to an anonymous referee for raising both of these worries.

degree. Similarly, an agent can have a high degree of confidence in the conjunction *p and I don't know p*, and yet this is maximally unassertable.[27] For those who would reduce assertability to credence, one option would be to modify the semantics above so that epistemic contradictions receive a probability of 0. For example, one option would be to rely on a dynamic semantics for conjunction, where the second conjunct is evaluated relative to the information state that results from updating with the first conjunct (see Groenendijk et al. 1996). This theory of conjunction predicts that $p \wedge \Diamond \neg p$ is false at every path, and so has a probability of 0.[28]

9 Fitelson's Resilient Equation

In this section, we link our results to a recent discussion of **Stalnaker's Thesis** by Branden Fitelson. Fitelson (2015) discusses a principle that he calls the "Resilient Equation":

> **Resilient Equation (RE).** For any descriptive A, B, C:
>
> $$Pr(\mathsf{A} \to \mathsf{B} \mid \mathsf{C}) = Pr(\mathsf{B} \mid \mathsf{A} \wedge \mathsf{C}).$$

As Fitelson shows, **RE** leads immediately to triviality, without need of appealing to the combination of **Closure**, **Ratio**, and **Conditionalization**. Fitelson takes this to show that RE should unequivocally be rejected. He traces the difficulty to a principle of conditional logic, Import-Export (that $\mathsf{A} \to (\mathsf{B} \to \mathsf{C}) \dashv\models (\mathsf{A} \wedge \mathsf{B}) \to \mathsf{C}$), which entails RE in combination with ST.

We agree that RE should be rejected—as Fitelson points out, the route from RE to triviality is inescapable. But we disagree with Fitelson's diagnosis. The problem does not have to do with Import-Export, but rather with the identification of updated probabilities and conditional probabilities, i.e. with the step that we rejected in §3. To better see the point, it's helpful to notice that the same scenarios that generated counterexamples to **Stalnaker's Thesis** and **Conditionalization** also generate coutnerexamples to **RE**. Consider the die example from above. Let A, B, and C be the following:[29]

> A: The die didn't land on two or four
>
> B: The die landed on six
>
> C: The die landed even

With this choice of A, B, and C, the relevant instance of **RE** is:

[27] For a recent account of degrees of assertability in terms of normality rather than credence, see Carter 2020.

[28] See Mandelkern 2019 for a variant of this idea that deals with other nearby kinds of epistemic contradictions.

[29] A similar example is discussed in Khoo and Santorio 2018

(29) $Pr(\neg(2 \text{ or } 4) > 6 \mid \text{even}) = Pr(\text{six} \mid \text{even} \wedge \neg\,(2 \text{ or } 4))$

Now, the right-hand side of RE goes to 1.

$$Pr(\mathsf{B} \mid \mathsf{A} \wedge \mathsf{C}) = Pr(\text{six} \mid \text{even} \wedge \neg\,(2 \text{ or } 4)) = Pr(\text{six} \mid \text{six}) = 1$$

However, the left-hand side is different from 1—via the calculation we went through in §3.2, we know that it's 1/2. So the equation in (29) fails, and we have a counterexample to **RE**.

What goes wrong here? Intuitively, something is wrong with (29). After learning that the die landed even, one should not assign credence 1/2 to the conditional *If the die didn't land on two or four, it landed on six*. Rather, that credence should be 1.

This diagnosis is exactly right. **RE** is motivated by a correct informal intuition: one's credence in $\mathsf{A} \to \mathsf{B}$, updated on C, should be identical to one's credence in C, updated on $\mathsf{A} \wedge \mathsf{C}$. But this intuition is formalized incorrectly if we take updated probabilities to be identical to conditional probabilities. Conversely, we obtain a tenable claim if we take update to work via **Hyperconditionalization**. The following is a simple consequence of the **Update Thesis**:

$$Pr_{\mathsf{C}}(\mathsf{A} \to \mathsf{B}) = Pr_{\mathsf{A} \wedge \mathsf{C}}(\mathsf{B})$$

Crucially, this claim drops the identification of updated probabilities with conditional probabilities. Once more, the lesson of triviality is that updated probabilities are not appropriately characterized, in general, via the ratio formula.

10 Conclusion

We began this paper by offering a unitary diagnosis of a cluster of triviality results about conditionals and epistemic modals. Conditionalization is not the rule of rational update when epistemic modalities and conditionals are involved. Second, building on existing work, we have developed a package of semantics and probability that vindicates several intuitive bridge principles linking epistemic modalities and probability. Of course, some issues about epistemic modality and probability are still outstanding. But, if we are right, we solve some important ones.

Our work is indebted to a number of similar theories which, like ours, exploit sequences of worlds. Yet the theory we propose is innovative both technically and conceptually. On the technical side: (i) our theory merges perfectly with standard compositional semantics for modals and conditionals; (ii) we vindicate a large number of bridge principles between probability and modality, including a version of Stalnaker's Thesis that is robust under update. On the conceptual side: we suggest that a core group of triviality results are due to the fact that Conditionalization fails when epistemic modality is involved. As informational theorists have pointed out, epistemic modalities and conditionals have a nonclassical logic. This nonclassical logic should be paired with a nonclassical theory of update.[30]

[30]Thanks to Fabrizio Cariani, Branden Fitelson, Melissa Fusco, Arc Kocurek, Justin Khoo, Kurt Norlin,

Appendix

In this appendix, we prove the theorems from the main body of the paper.

Theorem 1. For all C, for all descriptive A and B, and for any C that is the credence function of a well-behaved epistemic state E:

$$C(\mathsf{A} \to \mathsf{B}) = C(\mathsf{B} \mid \mathsf{A})$$

Proof. The proof is very similar to ideas in Khoo and Santorio 2018, with a few changes. When A and B are descriptive, $p \in [\![\mathsf{A} \to \mathsf{B}]\!]$ just in case B is true at the first A world in p. We calculate the probability of $\mathsf{A} \to \mathsf{B}$ by applying the Law of Total Probability across the partition of paths into cells that agree on where the first A world occurs.

Let P^n_{A} be the set of paths where A first occurs at the nth position. Then, where n is the cardinality of W, we have:

(30) $$C(\mathsf{A} \to \mathsf{B}) = C(P^1_{\mathsf{A}}) \times C(\mathsf{A} \to \mathsf{B} \mid P^1_{\mathsf{A}}) + \cdots + C(P^n_{\mathsf{A}}) \times C(\mathsf{A} \to \mathsf{B} \mid P^n_{\mathsf{A}})$$

Our strategy is to prove that for any i, we have $C(\mathsf{A} \to \mathsf{B} \mid P^i_{\mathsf{A}}) = Pr(\mathsf{B} \mid \mathsf{A})$. Then the overall probability of $\mathsf{A} \to \mathsf{B}$ is just $Pr(\mathsf{B} \mid \mathsf{A})$, since the weights on each individual term add up to 1.

To see why $C(\mathsf{A} \to \mathsf{B} \mid P^1_{\mathsf{A}}) = Pr(\mathsf{B} \mid \mathsf{A})$, first note that the conjunction $\mathsf{A} \to \mathsf{B} \wedge P^1_{\mathsf{A}}$ is the set of paths where the first A world occurs at p_1, and where $\mathsf{A} \to \mathsf{B}$ is true. These are just the paths where $\mathsf{A} \wedge \mathsf{B}$ is true at p_1, and so is simply $\mathsf{A} \wedge \mathsf{B}$. Thus we have:

$$\begin{aligned} C(\mathsf{A} \to \mathsf{B} \mid P^1_{\mathsf{A}}) &= \frac{C(\mathsf{A} \to \mathsf{B} \wedge P^1_{\mathsf{A}})}{C(P^1_{\mathsf{A}})} \\ &= \frac{C(\mathsf{A} \wedge \mathsf{B})}{C(\mathsf{A})} \\ &= \frac{Pr(\mathsf{A} \wedge \mathsf{B})}{Pr(\mathsf{A})} \\ &= Pr(\mathsf{B} \mid \mathsf{A}) \end{aligned} \tag{1}$$

Now we generalize this fact to arbitrary i. Here, there are two important steps. First, the conjunction $\mathsf{A} \to \mathsf{B} \wedge P^i_{\mathsf{A}}$ is the set of paths where the first A world occurs at p_i, and where $\mathsf{A} \to \mathsf{B}$ is true. As before, this is just the set of paths where the first A world occurs at p_i, and B is true there. Any such path consists of $i-1$ worlds where $\neg\mathsf{A}$ is true, and then a world where $\mathsf{A} \wedge \mathsf{B}$ is true.

The second step is to calculate the probability of P^i_{A} and of $\mathsf{A} \to \mathsf{B} \wedge P^i_{\mathsf{A}}$. The former set contains just those paths that consist of $i-1$ worlds where A is false, followed by a world where A is true. We can partition this set based on which worlds exactly make up the path up to i, so that:

Ginger Schultheis, Robbie Williams, and the Melbourne University Logic Seminar.

$$C(P^i_{\mathsf{A}}) = \sum_{w_{1,\dots,i-1} \in \neg\mathbf{A}} \sum_{w_i \in \mathbf{A}} C(p[w_1, \dots, w_{i-1}, w_i])$$

We can then understand this last quantity in terms of Pr by appeal to the construction of path probability, reaching:

$$C(P^i_{\mathsf{A}}) = \sum_{w_{1,\dots,i-1} \in \neg\mathbf{A}} \sum_{w_i \in \mathbf{A}} C(p[w_1, \dots, w_{i-1}]) \times \frac{Pr(w_i)}{Pr(W - \{w_1, \dots, w_{i-1}\})}$$

By similar reasoning, we have:

$$C(\mathsf{A} \to \mathsf{B} \wedge P^i_{\mathsf{A}}) = \sum_{w_{1,\dots,i-1} \in \neg\mathbf{A}} \sum_{w_i \in \mathbf{A} \wedge \mathbf{B}} C(p[w_1, \dots, w_i])$$

Again, we can reduce this quantity using the construction of path probability:

$$C(\mathsf{A} \to \mathsf{B} \wedge P^i_{\mathsf{A}}) = \sum_{w_{1,\dots,i-1} \in \neg\mathbf{A}} \sum_{w_i \in \mathbf{A} \wedge \mathbf{B}} C(p[w_1, \dots, w_{i-1}]) \times \frac{Pr(w_i)}{Pr(W - \{w_1, \dots, w_{i-1}\})}$$

All that is left is to solve for $\frac{C(\mathsf{A} \to \mathsf{B} \wedge P^i_{\mathsf{A}})}{C(P^i_{\mathsf{A}})}$. Here, most terms immediately cancel out, producing:

$$\frac{C(\mathsf{A} \to \mathsf{B} \wedge P^i_{\mathsf{A}})}{C(P^i_{\mathsf{A}})} = \sum_{w \in \mathbf{A} \wedge \mathbf{B}} \sum_{w' \in \mathbf{A}} \frac{Pr(w)}{Pr(w')} = Pr(\mathsf{B} \mid \mathsf{A})$$

The result is that $C(\mathsf{A} \to \mathsf{B} \mid P^i_{\mathsf{A}})$ is guaranteed to be $Pr(\mathsf{B} \mid \mathsf{A})$, no matter the choice of i. All that is left is to return to our initial application of the Law of Total Probability:

(31) a. $C(\mathsf{A} \to \mathsf{B}) = C(P^1_{\mathsf{A}}) \times C(\mathsf{A} \to \mathsf{B} \mid P^1_{\mathsf{A}}) + \dots + C(P^n_{\mathsf{A}}) \times C(\mathsf{A} \to \mathsf{B} \mid P^n_{\mathsf{A}})$

b. $C(\mathsf{A} \to \mathsf{B}) = C(P^1_{\mathsf{A}}) \times Pr(\mathsf{B} \mid \mathsf{A}) + \dots + C(P^n_{\mathsf{A}}) \times Pr(\mathsf{B} \mid \mathsf{A})$

c. $C(\mathsf{A} \to \mathsf{B}) = Pr(\mathsf{B} \mid \mathsf{A})$ $\square$

Theorem 2. For any A and for any C that is the credence function of a well-behaved epistemic state E:

1. $C(\Box \mathsf{A}) = \begin{cases} 1 & \text{if } C(\mathsf{A}) = 1 \\ 0 & \text{otherwise} \end{cases}$

2. $C(\Diamond \mathsf{A}) = \begin{cases} 1 & \text{if } C(\mathsf{A}) > 0 \\ 0 & \text{otherwise} \end{cases}$

Proof. First observe that all paths in a well behaved epistemic space contain the same live possible worlds. So either all the paths contain a live A world, or none do. In the former case, $\Diamond$A has a probability of 1; otherwise, it has a probability of 0. Similarly, in the former case $\Box\neg$A has a probability of 0; otherwise it has a probability of 1. All paths in a well behaved epistemic space contain a live A world just in case that space is based on an underlying probability function over worlds that assign some A world a positive probability. This holds just in case the epistemic space assigns a positive probability to A. □

Theorem 3. Let $E = \langle P, C\rangle$ be an epistemic state such that $E = \uparrow U$, with $U = \langle W, Pr\rangle$. Let $E^{\mathsf{A}} = \langle P^{\mathsf{A}}, C^{\mathsf{A}}\rangle = \uparrow U_{\mathsf{A}}$, where U_{A} is the proto-epistemic space $\langle W \cap [\![\mathsf{A}]\!], Pr_A\rangle$ we reach by conditionalizing U on A. Then:

$$C^{\mathsf{A}}(p) \quad = \quad \sum C(p') : p = p' + \mathsf{A} \quad = \quad C_{\mathsf{A}}(p)$$

Proof. Here is the overall structure of the proof. Our ultimate goal is to show that for any C and A, we have $C_{\mathsf{A}} = C^{\mathsf{A}}$. First, we show that for any C and A, there exists a descriptive proxy A^* such that $C_{\mathsf{A}} = C_{\mathsf{A}^*}$ and $C^{\mathsf{A}} = C^{\mathsf{A}^*}$. Second, we rely on the fact that C_{A^*} is equivalent to a series of updates of the form $C_{\neg w}$, and C^{A^*} is equivalent to a series of updates of the form $C^{\neg w}$. Finally, we show that $C_{\neg w} = C^{\neg w}$ for every w.

We begin by constructing the descriptive proxy A^* for any claim A and epistemic space E, where updating E with A is equivalent to updating E with A^*, no matter which update procedure one uses. To construct this proxy, we rely on the fact that updating an epistemic space with a claim supervenes on updating the paths in that space with A: if for every path in the space p, $p + \mathsf{A} = p + \mathsf{A}^*$, updating the overall space with A will be equivalent to updating with A^*.

Let's start by introducing a definition from Santorio 2019b. In §4, we defined a notion of path update: the update of a path p with A is the largest subsequence of p such that all of its permutations make A true. So far, we have not defined a notion of the update of an information state. Using $\textsc{path}(i)$ to denote the set of paths generated from i, we can define it as follows:

Information state update
$i + \mathsf{A} = \{w : w \in i$ and $\exists p \in \textsc{path}(i)$: w is a member of $p + \mathsf{A}\}$

The update of i with A is the set of worlds that appear in some path generated by i.

Now, we can prove that the procedure for updating paths and the procedure for updating information states are commutative, in the following sense:

Lemma 1. For all sets of paths P and all information states i such that $P = \textsc{path}(i)$, and for all A:

$$P_{\mathsf{A}} = \{p + \mathsf{A} : \ p \in P\} \quad = \quad \textsc{path}(i + \mathsf{A})$$

I.e.: if we take a set of paths and update them with A, or if we take the information state i that generates them, update i with A, and then generate a set of paths from the updated information state, we get the same results.

Proof. P is generated from i, so P includes all and only the sequences of worlds from i. Moreover, given the definition of path update, we know that, when we update each path $p \in P$ with A, we always remove from p the same worlds. It follows that the paths in P_{A} are all and only the sequences generated from some set of worlds S. So, to prove Lemma 1, all we need to show is that S is identical to $i + \mathsf{A}$. For *reductio*, suppose that this is not the case. Then (at least) one of the following two cases holds:

(i) for some world w, $w \in S$ but $w \notin i + \mathsf{A}$;

(ii) for some world w, $w \notin S$ but $w \in i + \mathsf{A}$

Assume (i). Then there is a world that is a member of all paths in S, but not of $i + \mathsf{A}$. But, by the definition of information state update, $w \in i + \mathsf{A}$ iff, for some path p in $\textsc{path}(i)$, w is a member of some path $p + \mathsf{A}$. Contradiction. Now assume (ii). Again by the definition of information state update, $w \in i + \mathsf{A}$ iff, for some path p in $\textsc{path}(i)$, w is a member of some path $p + \mathsf{A}$. Then, for some path $p + \mathsf{A}$ in P_{A}, w is a member of $p + \mathsf{A}$, and since all paths in P_{A} are composed of the same worlds, w is in S after all. Contradiction. □

The next step is based on a simple observation: both path and information state update can be modeled as a (possibly vacuous) operation of removal of worlds from a path or an information state. In particular, as pointed out in §4, updates with modalized sentences and conditionals can also be modeled in this way. Hence we have:

Lemma 2. For all i, all A, and all $P = \textsc{path}(i)$: there is a descriptive A^* such that $P_{\mathsf{A}} = P_{\mathsf{A}^*}$.

(A full proof would proceed by induction on the semantic clauses of §4, and is left to the reader.)

The next lemma follows immediately from Lemmas 1 and 2:

Lemma 3. For all i, all A, and all $P = \textsc{path}(i)$: there is a descriptive A^* such that:

$$\textsc{path}(i + \mathsf{A}) = P_{\mathsf{A}} = P_{\mathsf{A}^*} = \textsc{path}(i + \mathsf{A}^*)$$

Let us take stock. Given Lemma 3, we have established the following about the non-credal part of update in the system: every update in the system can be thought of as update with descriptive sentences. In addition, we have that the following two procedures yield the same result: we update paths directly, or we update the information state that generates a set of paths, and then generate a new set of paths. This establishes something like a qualitative counterpart of Theorem 3. At this point, we can move to considering the way that credences are updated.

Take an epistemic space E lifted from proto-epistemic space U, containing P and W respectively. Lemma 3 guarantees that when we update E with A, we have a descriptive proxy A^* where $C_A = C_{A^*}$ and $C^A = C^{A^*}$. In particular, where P is the set of paths in C, the descriptive proxy is the set of worlds in $P + A$.

To complete the proof of Theorem 3, we show that $C_{A^*} = C^{A^*}$. Here, the first observation is that A^* is equivalent to a conjunction of claims of the form $\neg w$, for each world w where A^* is false. (Crucially, since A^* is descriptive, this notion is well-defined.) In addition, we know that both Hyperconditionalization and our other update procedure are commutative for descriptive claims. Updating on a conjunction of claims of the form $\neg w$ is equivalent to updating on each claim of the form $\neg w$ in order. In the case of hyperconditionalization, this is because the underlying notion of path updating is commutative for descriptive claims: $(p + \mathsf{A}) + \mathsf{B} = p + (\mathsf{A} \wedge \mathsf{B})$. In the case of our second notion of updating, this is because the underlying notion of conditionalization is commutative for descriptive claims: $(W + \mathsf{A}) + \mathsf{B} = W + (\mathsf{A} \wedge \mathsf{B})$. To complete our proof, then, all that we must show is that for any claim $\neg w$, $C_{\neg w} = C^{\neg w}$.

We now show that Theorem 3 holds when A is restricted to propositions of the form $\neg w$, for some choice of world w. Theorem 3 follows quickly from this lemma, since we can express any descriptive claim A as a conjunction of propositions of the form $\neg w$, for each world w eliminated by A, and updating on any such conjunction using either update procedure is simply a matter of updating on each claim one at a time.

We now prove:

Lemma 4. Let $E = \langle P, C\rangle$ be an epistemic state such that $E = \uparrow U$, with $U = \langle W, Pr\rangle$. Let $E^{\neg w} = \langle P^{\neg w}, C^{\neg w}\rangle = \uparrow U_{\neg w}$, where $U_{\neg w}$ is the proto-epistemic space we reach by conditionalizing U on $\neg w$. We have:

$$C^{\neg w}(p) = \sum C(p') : p = p' + \neg w = C_{\neg w}(p)$$

We prove this lemma by considering a series of worlds 1 through n, and arbitrarily instantiating w with the value n. For readability, we are going to denote worlds by using natural numbers throughout the proof.

The proof works by comparing two values, $C_{\neg n}(\langle 1, \ldots, n-1\rangle)$ and $C^{\neg n}(\langle 1, \ldots, n-1\rangle)$, which result from updating $C(\langle 1, \ldots, n-1, n\rangle)$ in two different ways. $C(\langle 1, \ldots, n-1, n\rangle)$, recall, is the probability of the path $\langle 1, \ldots, n-1, n\rangle$. The first of the two values, $C_{\neg n}(\langle 1, \ldots, n-1\rangle)$, is what we get by hyperconditionalizing on the information $\{1, \ldots, n-i\}$.[31] The second of the two values, $C^{\neg n}(\langle 1, \ldots, n-1\rangle)$, is what we get by first updating the proto-epistemic state U via simple conditionalization on the information $\{1, \ldots, n-i\}$ and then re-lifting the updated proto-epistemic state. Throughout the proof, this particular choice of path is arbitrary; the proof could be repeated for any path that involves the same worlds, arranged in a different sequence.

To start, we notice that, by the definition of Hyperconditionalization, the value of $C_{\neg n}(\langle 1, \ldots, n-1\rangle)$ equals the sum of the values of all the probabilities of the paths that we get by inserting n in $\langle 1, \ldots, n-1\rangle$ in any position. In short:

[31] There is a slight abuse of terminology here. Strictly speaking, hyperconditionalization updates paths with sentences, not propositions. We can understand update in terms of propositions as an update produced by any descriptive sentence that expresses the relevant proposition.

(32) $C_{\neg n}(\langle 1,\ldots,n-1\rangle) = C(\langle 1,\ldots,n\rangle) + C(\langle n,1,\ldots,n-1\rangle) + C(\langle 1,n,\ldots,n-1\rangle) + \cdots + C(\langle 1,\ldots,n,n-1\rangle)$

Our proof will proceed by showing that the probability value we get for $\langle 1,\ldots,n-1\rangle$ by conditionalizing on the proto-epistemic space and re-lifting equals the right-hand side of (32). I.e., we show:

(33) $C^{\neg n}(\langle 1,\ldots,n-1\rangle) = C(\langle 1,\ldots,n\rangle) + C(\langle n,1,\ldots,n-1\rangle) + C(\langle 1,n,\ldots,n-1\rangle) + \cdots + C(\langle 1,\ldots,n,n-1\rangle)$

For readability, we let Pi denote the probability according to Pr of world i. We let Pij denote the probability of the disjunction of world i with world k.

First, we notice that, by the definition of path probability in §6, the way we calculate $C^{\neg n}(\langle 1,\ldots,n-1\rangle)$ is the following (where P is the probability function we get after conditionalization on $\{1,\ldots,n-1\}$):

(34) $C^{\neg n}(\langle 1,\ldots,n-1\rangle) = \frac{P1\times\cdots\times P(n-2)}{P1\ldots(n-1)\times P2\ldots(n-1)\times\cdots\times P(n-2)(n-1)}$

To prove Lemma 4, we must show that the fraction on the right-hand side of (34) is identical to the sum on the right-hand side of in (33), i.e. the sum in (35):

(35) $C(\langle 1,\ldots,n\rangle) + C(\langle n,1,\ldots,n-1\rangle) + C(\langle 1,n,\ldots,n-1\rangle) + \cdots + C(\langle 1,\ldots,n,n-1\rangle)$

To start, let us consider how the probability of each of the terms in (35) is calculated, according to the definition of path probability in §6:

path	C
$1,\ldots,n$	$\frac{P1\times\cdots\times Pn}{P2\ldots n\times P3\ldots n\times\cdots\times Pn}$
$n,1,\ldots,n-1$	$\frac{P1\times\cdots\times Pn}{P1\ldots(n-1)\times P2\ldots(n-1)\times\cdots\times P(n-1)}$
$1,n,2,\ldots,n-1$	$\frac{P1\times\cdots\times Pn}{P2\ldots n\times P2\ldots(n-1)\times\cdots\times P(n-1)}$
$\ldots$	$\ldots$
$1,\ldots,n,n-1$	$\frac{P1\times\cdots\times Pn}{P2\ldots n\times P3\ldots n\times\cdots\times P(n-1)n\times P(n-1)}$

We can prove the result we want by some simple algebraic manipulations. First, we factor out $P1\times\cdots\times Pn$ from the sum of the terms of (35), to reach:

(36) $P1\times\cdots\times Pn\times[\frac{1}{P2\ldots n\times P3\ldots n\times\cdots\times Pn} + \frac{1}{P1\ldots(n-1)\times P2\ldots(n-1)\times\cdots\times P(n-1)} + \frac{1}{P2\ldots n\times P2\ldots(n-1)\times\cdots\times P(n-1)} + \cdots + \frac{1}{P2\ldots n\times P3\ldots n\times\ldots P(n-1)n\times P(n-1)}]$

Now we divide the first term by $\frac{P(n-1)}{P(n-1)}$ and the remaining terms by $\frac{Pn}{Pn}$.

(37) $P1\times\cdots\times Pn\times[\frac{P(n-1)}{P2\ldots n\times P3\ldots n\times\cdots\times Pn\times P(n-1)} + \frac{Pn}{P1\ldots(n-1)\times P2\ldots(n-1)\times\cdots\times P(n-1)\times Pn} + \frac{Pn}{P2\ldots n\times P2\ldots(n-1)\times\cdots\times P(n-1)\times Pn} + \cdots + \frac{Pn}{P2\ldots n\times P3\ldots n\times\ldots P(n-1)n\times P(n-1)\times Pn}]$

Simplifying, we reach:

$$(38) \qquad P1\times\cdots\times P(n-2)\times[\frac{P(n-1)}{P2...n\times P3...n\times\cdots\times P(n-1)n}+\frac{Pn}{P1...(n-1)\times P2...(n-1)\times\cdots\times P(n-2)(n-1)}+\frac{Pn}{P2...n\times P2...(n-1)\times\cdots\times P(n-2)(n-1)}+\cdots+\frac{Pn}{P2...n\times P3...n\times...P(n-1)n}]$$

Now we can simplify this in stages, first by summing the first and last of these terms. Here, the key is that $P(n-1)n$ is equal to $P(n-1) + Pn$. When the first and last term are summed, we can therefore eliminate the numerator and the term $P(n-1)n$ from the denominator:

$$(39) \qquad P1\times\cdots\times P(n-2)\times[\frac{1}{P2...n\times P3...n\times\cdots\times P(n-2)...n}+\frac{Pn}{P1...(n-1)\times P2...(n-1)\times\cdots\times P(n-2)(n-1)}+\frac{Pn}{P2...n\times P2...(n-1)\times\cdots\times P(n-2)(n-1)}+\cdots+\frac{Pn}{P2...n\times P3...n\times\cdots\times P(n-2)...n\times P(n-2)(n-1)}]$$

Now we multiply the first term by $\frac{P(n-2)(n-1)}{P(n-2)(n-1)}$ so that it has the same denumerator as the last term:

$$(40) \qquad P1\times\cdots\times P(n-2)\times[\frac{P(n-2)(n-1)}{P2...n\times P3...n\times\cdots\times P(n-2)...n\times P(n-2)(n-1)}+\frac{Pn}{P1...(n-1)\times P2...(n-1)\times\cdots\times P(n-2)(n-1)}+\frac{Pn}{P2...n\times P2...(n-1)\times\cdots\times P(n-2)(n-1)}+\cdots+\frac{Pn}{P2...n\times P3...n\times\cdots\times P(n-2)...n\times P(n-2)(n-1)}]$$

Now we add the first and last term:

$$(41) \qquad P1\times\cdots\times P(n-2)\times[\frac{P(n-2)(n-1)+Pn}{P2...n\times P3...n\times\cdots\times P(n-2)...n\times P(n-2)(n-1)}+\frac{Pn}{P1...(n-1)\times P2...(n-1)\times\cdots\times P(n-2)(n-1)}+\frac{Pn}{P2...n\times P2...(n-1)\times\cdots\times P(n-2)(n-1)}+\cdots+\frac{Pn}{P2...n\times P(n-3)n\times P(n-2)(n-1)}]$$

Applying additivity, we know that $P(n-2)(n-1)n = P(n-2)(n-1) + Pn$, so we can simplify:

$$(42) \qquad P1\times\cdots\times P(n-2)\times[\frac{1}{P2...n\times P3...n\times\cdots\times P(n-3)...n\times P(n-2)(n-1)}+\frac{Pn}{P1...(n-1)\times P2...(n-1)\times\cdots\times P(n-2)(n-1)}+\frac{Pn}{P2...n\times P2...(n-1)\times\cdots\times P(n-2)(n-1)}+\cdots+\frac{Pn}{P2...n\times P(n-3)n\times P(n-2)(n-1)}]$$

We continue this process of converting the first term to have the same denumerator as the last term and adding the result. Throughout this process, the number of terms continues to shrink from right to left until the only unmanipulated term is $C(\langle n, 1, \ldots, (n-1)\rangle)$:

$$(43) \qquad P1\times\cdots\times P(n-2)\times[\frac{1}{P2...(n-1)\times P2...(n-1)\times\cdots\times P(n-2)(n-1)}+\frac{Pn}{P1...(n-1)\times P2...(n-1)\times\cdots\times P(n-2)(n-1)}]$$

Multiplying the first term by $\frac{P1(n-1)}{P1(n-1)}$, we reach:

$$(44) \qquad P1\times\cdots\times P(n-2)\times[\frac{P1(n-1)}{P1(n-1)\times P2...(n-1)\times P2...(n-1)\times\cdots\times P(n-2)(n-1)}+\frac{Pn}{P1...(n-1)\times P2...(n-1)\times\cdots\times P(n-2)(n-1)}]$$

Adding we have:

$$(45) \qquad P1\times\cdots\times P(n-2)\times[\frac{P1(n-1)+Pn}{P1(n-1)\times P2...(n-1)\times P2...(n-1)\times\cdots\times P(n-2)(n-1)}]$$

Additivity implies that $P1(n-1) + Pn = 1$. This gives us:

(46) $C(\langle 1,\ldots,n\rangle) + C(\langle n,1,\ldots,n-1\rangle) + C(\langle 1,n,\ldots,n-1\rangle) + \cdots + C(\langle 1,\ldots,n,n-1\rangle) = \frac{P1\times\cdots\times P(n-2)}{P(1\ldots(n-1))\times P(2\ldots(n-1))\times\cdots\times P((n-2)(n-1))} = C_{\neg n}(\langle 1,\ldots,n-1\rangle)$ □

This completes the proof of Theorem 3.

Theorem 4. Suppose E is well-behaved. Then, for all descriptive A, $E +_{\mathsf{HC}} \mathsf{A}$ is well-behaved.

Proof. Theorem 4 follows immediately from Theorem 3. Since the result of hyperconditionalizing C on A, C_{A}, is equivalent to C^{A}, and since the latter is a well-behaved credence distribution by construction, it immediately follows that hyperconditionalizing on any sentence leads to a well-behaved information state. □

References

Ernest Adams. *A Primer of Probability Logic*. CSLI Publications, Stanford, 1998.

Ernest Wilcox Adams. *The logic of conditionals: An application of probability to deductive logic*, volume 86. Springer Science & Business Media, 1975.

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

David Beaver. *Presupposition and Assertion in Dynamic Semantics*. CSLI Publications, Stanford, CA, 2001.

Bob Beddor and Simon Goldstein. Believing epistemic contradictions. *Review of Symbolic Logic*, 11(1), March 2018.

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

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

Richard Bradley. Multidimensional possible-world semantics for conditionals. *Philosophical Review*, 4:539–571, 2012.

Sam Carter. Degrees of assertability. *Philosophy and Phenomenological Research*, 2020.

Nate Charlow. Triviality for restrictor conditionals. *Noûs*, 2015.

Nate Charlow. Grading modal judgment. *Mind*, Forthcoming.

Alexandre Cremers, Manuel Križ, and Emmanuel Chemla. *Probability Judgments of Gappy Sentences*, pages 111–150. Springer International Publishing, 2017.

Kenny Easwaran. Regularity and hyperreal credences. *Philosophical Review*, 123(1): 1–41, 2014.

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

Jonathan St. B. T. Evans and David E. Over. *If*. Oxford University Press, Oxford, 2004.

Branden Fitelson. The strongest possible lewisian triviality result. *Thought: A Journal of Philosophy*, 2015.

Peter Gardenfors. Imaging and conditionalization. *The Journal of Philosophy*, 79(12): 747–760, 1982.

Anthony S. Gillies. Epistemic conditionals and conditional epistemics. *Nous*, 38(4): 585–616, 2004.

Simon Goldstein. Triviality results for probabilistic modals. *Philosophy and Phenomenological Research*, December 2017.

Simon Goldstein. Generalized update semantics. *Mind*, 2018.

Simon Goldstein. Free choice impossibility results. *Journal of Philosophical Logic*, (49): 249–282, 2020.

Jeroen Groenendijk, Martin Stokhof, and Frank Veltman. Coreference and modality. In Shalom Lappin, editor, *Handbook of Contemporary Semantic Theory*, pages 179–213. Blackwell, 1996.

Alan Hájek. What conditional probability could not be. *Synthese*, 137(3):273–323, 2003. doi: 10.1023/b:synt.0000004904.91112.16.

Alan Hájek and N. Hall. The Hypothesis of the Conditional Construal of Conditional Probability. In Ellery Eells, Brian Skyrms, and Ernest W. Adams, editors, *Probability and Conditionals: Belief Revision and Rational Decision*, page 75. Cambridge University Press, 1994.

Richard Jeffrey and Robert Stalnaker. Conditionals as random variables. In Ellery Eells, Brian Skyrms, and Ernest W. Adams, editors, *Probability and Conditionals: Belief Revision and Rational Decision*, page 31. Cambridge University Press, 1994.

Magdalena Kaufmann and Stefan Kaufmann. *The Handbook of Contemporary Semantic Theory*, chapter Conditionals and Modality. Number 8. Wiley and Sons, 2015.

Stefan Kaufmann. Conditioning against the grain. *Journal of Philosophical Logic*, 33 (6):583–606, 2004. doi: 10.1023/b:logi.0000046142.51136.bf.

Stefan Kaufmann. Conditionals right and left: Probabilities for the whole family. *Journal of Philosophical Logic*, 38(1):1–53, 2009.

Stefan Kaufmann. Conditionals, conditional probabilities, and conditionalization. In Hans-Christian Schmitz and Henk Zeevat, editors, *Bayesian Natural Language Semantics and Pragmatics*, pages 71–94. springer, 2015.

Justin Khoo. A note on gibbard's proof. *Philosophical Studies*, 166(S1):153–164, 2013. doi: 10.1007/s11098-012-0062-x.

Justin Khoo. *The Meaning of* If. Oxford University Press, forthcoming.

Justin Khoo and Paolo Santorio. Lecture notes: Probability of conditionals in modal semantics, 2018. Lecture notes available at `http://www.justinkhoo.org/KS-NASSLLI2018.pdf`).

Nathan Klinedinst and Daniel Rothschild. Connectives without truth tables. *Natural Language Semantics*, 20(2):137–175, June 2012.

Angelika Kratzer. *Modals and Conditionals: New and Revised Perspectives*, volume 36. Oxford University Press, 2012.

David Lewis. Probabilities of conditionals and conditional probabilities. *The Philosophical Review*, 85:297–315, 1976.

Matt Mandelkern. Bounded modality. *Philosophical Review*, 2019.

Vann McGee. A counterexample to modus ponens. *The Journal of Philosophy*, 82(9): 462–471, 1985.

Vann McGee. Conditional probabilities and compounds of conditionals. *Philosophical Review*, 98(4):485–541, 1989. doi: 10.2307/2185116.

Sarah Moss. On the semantics and pragmatics of epistemic modals. *Semantics and Pragmatics*, 2015.

Dilip Ninan. Relational semantics and domain semantics for epistemic modals. *Journal of Philosophical Logic*, 2016.

Kurt Norlin. In the logic of certainty, $\supset$ is stronger than $\rightarrow$. *Thought: A Journal of Philosophy*, 1(9):58–63, 2020.

Kurt Norlin. A stalnaker semantics for mcgee conditionals. *Erkenntnis*, 1(86):59–70, 2021.

Frank Ramsey. Truth and probability. In Antony Eagle, editor, *Philosophy of Probability: Contemporary Readings*, pages 52–94. Routledge, 1926.

Jeff Russell and John Hawthorne. General dynamic triviality theorems. *Philosophical Review*, 125(3):307–339, 2016.

Paolo Santorio. Conditional excluded middle in informational semantics. In *Proceedings of the Amsterdam Colloquium*, volume 21, 2018.

Paolo Santorio. Trivializing informational consequence. Draft, University of Maryland at College Park, 2019a.

Paolo Santorio. Path semantics for indicative conditionals. forthcoming in *Mind*, 2019b.

Paolo Santorio and J.R.G. Williams. Triviality and indeterminacy. Draft, University of Maryland, College Park, and University of Leeds.

Barry Schein. Adverbial, descriptive reciprocals. *Philosophical Perspectives*, 17(1):333–367, 2003.

Philippe Schlenker. Conditionals as definite descriptions. *Research on language and computation*, 2(3):417–462, 2004.

Moritz Schulz. Epistemic modals and informational consequence. *Synthese*, 174(3): 385–395, 2010.

Robert Stalnaker. A theory of conditionals. *American Philosophical Quarterly*, pages 98–112, 1968.

Robert Stalnaker. A defense of conditional excluded middle. In William L. Harper, Robert Stalnaker, and Glenn Pearce, editors, *IFS: Conditionals, Belief, Decision, Chance and Time*, volume 15 of *The University of Western Ontario Series in Philosophy of Science*, pages 87–104. Springer Netherlands, 1981.

Robert Stalnaker. *Inquiry*. MIT Press, 1984.

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

Eric Swanson. How not to theorize about the language of subjective uncertainty. In Andy Egan and B. Weatherson, editors, *Epistemic Modality*. Oxford University Press, 2011.

Eric Swanson. The application of constraint semantics to the language of subjective uncertainty. *Journal of Philosophical Logic*, pages 1–28, 2012.

Bas van Fraassen. Probabilities of conditionals. In C. Hooker, editor, *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*. 1976.

Frank Veltman. *Logics for Conditionals*. PhD thesis, University of Amsterdam, 1985.

Frank Veltman. Defaults in update semantics. *Journal of Philosophical Logic*, 25(3): 221–261, 1996.

Kai von Fintel and Anthony S. Gillies. An opinionated guide to epistemic modality. In Tamar Gendler and John Hawthorne, editors, *Oxford Studies in Epistemology*, volume 2, pages 32–62. Tamar Gendler and John Hawthorne, Oxford, oxford university press edition, 2007.

Malte Willer. Dynamics of epistemic modality. *Philosophical Review*, 122(1):44–92, 2013.

Malte Willer. Simplifying counterfactuals. In *Proceedings of the 20th Amsterdam Colloquium*, pages 428–437, 2015.

J. Robert G. Williams. Defending conditional excluded middle. *Nous*, 44(4):650–668, 2010.

J. Robert G. Williams. Counterfactual triviality. *Philosophy and Phenomenological Research*, 85(3):648–670, 2012.

Timothy Williamson. *Knowledge and its Limits*. Oxford University Press, Oxford, 2000.

Seth Yalcin. Epistemic modals. *Mind*, 116(464):983–1026, 2007.

Seth Yalcin. Epistemic modality *De Re*. *Ergo*, 2(19), 2015.