

Conditionals, Conditional Probabilities, and Conditionalization^{*}

Stefan Kaufmann

University of Connecticut
stefan.kaufmann@uconn.edu

Abstract. Philosophers investigating the interpretation and use of conditional sentences have long been intrigued by the intuitive correspondence between the probability of a conditional ‘*if A, then C*’ and the conditional probability of *C*, given *A*. Attempts to account for this intuition within a general probabilistic theory of belief, meaning and use have been plagued by a danger of trivialization, which has proven to be remarkably recalcitrant and absorbed much of the creative effort in the area. But there is a strategy for avoiding triviality that has been known for almost as long as the triviality results themselves. What is lacking is a straightforward integration of this approach in a larger framework of belief representation and dynamics. This paper discusses some of the issues involved and proposes an account of belief update by conditionalization.

1 Introduction

Most contemporary theories of the interpretation of conditionals are inspired by Ramsey’s (1929) paraphrase of the mental process involved in their interpretation, known as the **Ramsey Test (RT)**:

- (RT) If two people are arguing ‘*If A will C?*’ and are both in doubt as to *A*, they are adding *A* hypothetically to their stock of knowledge and arguing on that basis about *C*... We can say they are fixing their degrees of belief in *C* given *A*.

For all its intuitive appeal, (RT) is too general and underspecified to be operationalized in a concrete theoretical approach. To turn it into a definition, one has to flesh out several abstract ideas it mentions. Most crucially, this concerns the notions of **stock of knowledge**, **adding *A* temporarily**, and **degrees of belief**.

^{*} Thanks to Hans-Christian Schmitz and Henk Zeevat for organizing the ESSLLI 2014 workshop on Bayesian Natural-Language Semantics and Pragmatics, where I presented an early version of this paper. Other venues at which I presented related work include the research group “What if” at the University of Konstanz and the Logic Group at the University of Connecticut. I am grateful to the audiences at all these events for stimulating discussion and feedback. Thanks also to Hans-Christian Schmitz and Henk Zeevat for their extreme patience during the preparation of this manuscript. All errors and misrepresentations are my own.

In this paper I explore one particular way to read (RT), *viz.* a version of probabilistic semantics inspired by Ramsey and developed by Jeffrey (1964), Adams (1965, 1975), and many subsequent authors. The basic ingredients are familiar from propositional logic. Sentences denote **propositions**, modeled as sets of **possible worlds**. Epistemic states (Ramsey’s “stocks of knowledge”) are represented in terms of (subjective) **probabilities**. The addition of the antecedent proceeds by **conditionalization**.

Regarding the latter point, another question that needs to be addressed is whether one and the same update operation is appropriate for all conditionals. Conditionalization on a proposition is typically interpreted as modeling the process of **learning** (hence coming to believe) that the proposition is true. However, there are well-known examples of conditionals in which the relevant supposition is intuitively that the antecedent is **true** without its truth being **known** to the speaker. This can be illustrated with examples whose antecedents or consequents explicitly deny the relevant belief, such as ‘*if it’s raining and I don’t know it. . .*’. Examples of this kind are generally problematic for theories – whether probabilistic or not – which rely on a single operation to represent the hypothetical reasoning involved. I return to this issue towards the end of the paper, having introduced a formal framework in which the relevant distinction can be accounted for.

2 Probability models

To get things started, I begin with a fairly simple model-theoretic interpretation. For now the goal is to stay close to the standard possible-worlds apparatus of propositional logic, adding only the structure required to model probability judgments. The setup I introduce in this subsection will ultimately not work, as we will see below. But it is a good place to start.

I assume that sentences of English are mapped to expressions of the standard language of propositional logic, and spell out the formal details in terms of the latter. Let \mathcal{A} be a set of atomic propositional letters, and \mathcal{L}_A^0 be the smallest set containing \mathcal{A} and closed under negation and conjunction. I write ‘ $\bar{\varphi}$ ’ and ‘ $\varphi\psi$ ’ for the negation of φ and the conjunction of φ and ψ , respectively. A **probability model** is a standard possible-worlds model for \mathcal{L}_A^0 , augmented with a probability measure.

Definition 1 (Probability model). *A probability model for language \mathcal{L}_A^0 is a tuple $\langle \Omega, \mathcal{F}, \text{Pr}, V \rangle$, where:¹*

- Ω is a non-empty set (of possible worlds);
- \mathcal{F} is a σ -algebra on Ω ;
- Pr is a probability measure on \mathcal{F} ;
- V is a function mapping sentences to characteristic functions of elements of \mathcal{F} , subject to the following conditions, for all $\varphi, \psi \in \mathcal{L}_A^0$ and $\omega \in \Omega$:

$$V(\bar{\varphi})(\omega) = 1 - V(\varphi)(\omega)$$

$$V(\varphi\psi)(\omega) = V(\varphi)(\omega) \times V(\psi)(\omega)$$

To relate the terminology familiar from possible-worlds semantics to the statistical jargon of probability theory, we can equate possible worlds with **outcomes** and sets of possible worlds (i.e., propositions) with **events**. Definition 1 follows the practice, common in probability theory, of separating the representations of the space of outcomes and the algebra on which the measure is defined (here Ω and \mathcal{F} , respectively). There are good philosophical reasons for doing this, but for the purposes of this paper, no harm is done by assuming for simplicity that \mathcal{F} is the powerset of Ω . It then follows without further stipulation that all propositions denoted by sentences are in the domain of the probability function.

Note that V maps sentences not to sets of worlds, but to their characteristic functions. Statistically speaking, those sentence denotations are **indicator variables** – a special kind of **random variables** whose range is restricted to the set $\{0, 1\}$. In the present context this perspective was first proposed by Jeffrey (1991) and further developed by Stalnaker and Jeffrey (1994). The motivation for modeling the denotations of sentences in this way will become clear below. It is important in the formal setup, but I will occasionally collapse talk of sets of worlds and talk of their characteristic functions where concepts rather than implementation are at stake.

The function Pr assigns probabilities to propositions, not sentences, but based on it we can assign probabilities to sentences indirectly as the **expectations** of their values. Generally speaking (and not just in the case of variables with range $\{0, 1\}$), the expectation of a random variable is the weighted sum of its values, where the weights are the probabilities that it takes those values.

Definition 2 (Expectation). *Let $\theta : \Omega \mapsto \mathbb{R}$ be a random variable. The **expectation** of θ relative to a probability function Pr is defined as follows:²*

$$E[\theta] := \sum_{x \in \text{range}(\theta)} x \times \text{Pr}(\theta = x)$$

For the purposes of interpreting the language \mathcal{L}_A^0 , the relevant variable for a given sentence φ is its interpretation $V(\varphi)$.

Definition 3 (Probabilities of sentences). *Given a probability model for \mathcal{L}_A^0 (see Definition 1), a function P maps sentences to real numbers as follows:*

$$P(\varphi) := E[V(\varphi)]$$

¹ A σ -algebra on Ω is a non-empty set of subsets of Ω that is closed under complements and countable unions. A probability measure on \mathcal{F} is a countably additive function from \mathcal{F} to the real interval $[0, 1]$ such that $\text{Pr}(\Omega) = 1$.

² I write ‘ $\text{Pr}(\theta = x)$ ’ to refer to the probability of the event that θ has value x . This is an abbreviation of the more cumbersome ‘ $\text{Pr}(\{\omega \in \Omega \mid \theta(\omega) = x\})$ ’. I also assume, here and throughout this paper, that the range of the random variable is finite. This is guaranteed for \mathcal{L}_A^0 under V in Definition 2, but becomes a non-trivial restriction in general. Nothing hinges on it, however: Giving it up would merely require that the summations in the definitions be replaced with integrals.

Since $V(\varphi)$ is an indicator variable for all φ in \mathcal{L}_A^0 , the expectation of $V(\varphi)$ is just the probability that φ is true:

$$(2) \quad \begin{aligned} P(\varphi) &= 0 \times \Pr(V(\varphi) = 0) + 1 \times \Pr(V(\varphi) = 1) \\ &= \Pr(V(\varphi) = 1) \end{aligned}$$

Thus in this framework we can state the connection between the probabilities of sentences on the one hand and the probabilities of propositions on the other, in the disquotational slogan in (3), the probabilistic analog of the famous Tarskian truth definition.

$$(3) \quad P(\text{“snow is white”}) = \Pr(\text{snow is white})$$

Despite the equality, however, it is important to keep in mind the formal distinction between \Pr , a probability measure on an algebra of propositions, and P , an assignment of probabilities to sentences. Only the former is technically a **measure** on its domain, whereas the latter maps sentences to probabilities only indirectly. We will see below that this separation of a sentence’s denotation from its probability is useful in extending the interpretation to conditionals.

3 Conditionals and conditional probability

While Ramsey’s formulation in (RT) is open to a variety of interpretations, the reading that is at issue here – one which many philosophers have found eminently plausible – is that the “degree of C given A ” that speakers report and aim to adjust when using indicative conditionals, is just the conditional probability of C given A . This idea has wide appeal among philosophers, going back to Jeffrey (1964), Adams (1965, 1975), and Stalnaker (1970), among others.

3.1 Conditional Expectation

In the present framework, where the probabilities of sentences are defined as the expectations of their truth values, the corresponding idea is that the probabilities of conditionals are the **conditional expectations** of their consequents, given that their antecedents are true. The conditional expectation of a random variable is its expectation relative to the conditionalized probability distribution:

Definition 4 (Conditional expectation). *Let θ, ξ be random variables and $y \in \text{range}(\xi)$. The **conditional expectation** of θ , given $\xi = y$, is defined as*

$$E[\theta|\xi = y] = \sum_{x \in \text{range}(\theta)} x \times \Pr(\theta = x|\xi = y)$$

This suggests that we could supplement the function P , which maps sentences in \mathcal{L}_A^0 to expectations, with a two-place function mapping **pairs** of sentences to **conditional** expectations:

Definition 5 (Conditional probabilities of sentences). *Given a model for \mathcal{L}_A^0 (see Definition 1), a two-place function $P(\cdot|\cdot)$ maps pairs of sentences to conditional expectations:*

$$P(\varphi|\psi) = E[V(\varphi)|V(\psi) = 1]$$

It is then easy to show, along the lines of (2) above, that $P(\varphi|\psi)$ is defined just in case the conditional probability $\Pr(V(\varphi) = 1|V(\psi) = 1)$ is, and that, where defined, the two coincide. In this sense, we could say that $P(\cdot|\cdot)$ satisfies the probabilistic reading of (RT).

It is also quite clear, however, that this approach falls far short of giving us what we want. For what we ultimately want is an assignment of probabilities to conditional sentences, rather than the pairs of sentences that constitute them. I take it to be self-evident that we should favor an approach which treats conditionals on a par with the atomic sentences and Boolean compounds that make up the rest of the language. But if there is any need for a further argument for this claim, suffice it to point out that conditionals can be embedded and compounded with each other and with other sentences. The following are all well-formed and interpretable sentences of ordinary English:

- (4) a. If this match is wet, it won't light if you strike it.
 b. If this switch will fail if it is submerged in water, it will be discarded.
 c. If this vase will crack if it is dropped on wood, it will shatter if it is dropped on marble.

Compounds like those in (4) have sometimes been claimed to bear only a superficial resemblance to compounded conditionals. Semantically, the story goes, their constituents are subjected to some kind of re-interpretation as sentences in \mathcal{L}_A^0 , so that what looks like complex conditionals in are in fact simple ones (Adams, 1975; Gibbard, 1981, among others). But we should be wary of this kind of story, which opens up a suspiciously handy escape hatch in the face of technical problems posed by these sentences for one's favorite theory. There certainly is no **linguistic** evidence that these sentences are anything other than compound conditionals. A theory which assigns probabilities to these sentences in a systematic fashion may still be found wanting on empirical grounds; but at least it would lend itself to empirical verification in the first place.

What, then, would it take to extend P to conditionals? So far they are not even represented in the language. \mathcal{L}_A^0 , being closed under the usual Boolean operations only, comes with the means to express the material conditional (definable as $\overline{\psi\varphi}$); but this is not the intended interpretation of the natural-language conditional 'if ψ then φ ', nor is its probability equivalent to the conditional probability of φ given ψ .³

³ This is not the place to rehearse the arguments for and against the material conditional as an adequate rendering of our intuitions about the meaning of the 'if-then' construction. The material analysis has its adherents in philosophy (Jackson, 1979; Lewis, 1986, among many others) and linguistics (see Abbott, 2004, for recent ar-

So let us first augment the language with a two-place propositional connective ‘ \rightarrow ’ as a stand-in for the English ‘*if-then*’ construction. Let \mathcal{L}_A be the smallest set containing \mathcal{L}_A^0 and closed under \rightarrow . We can now state precisely what (RT) and Definition 5 imply in this framework: The probability assigned to $\psi \rightarrow \phi$ should be equal to the conditional probability assigned to its constituents, as shown in (5a). By the definition of P, this means that the **unconditional** expectation of the conditional should equal the **conditional** expectation of the consequent, given that antecedent is true (5b):

- (5) a. $P(\psi \rightarrow \varphi) = P(\varphi|\psi)$
 b. $E[V(\psi \rightarrow \varphi)] = E[V(\varphi)|V(\psi) = 1]$

To be quite clear, (5) is not a definition. Rather, it states a desideratum that one would like to achieve when extending the domain of V and P from \mathcal{L}_A^0 to \mathcal{L}_A . Interestingly, this extension is not at all straightforward. I turn to this issue next.

3.2 Triviality

The unification of (RT) on its probabilistic reading and a truth-conditional interpretation of conditionals has been a challenge ever since Lewis (1976) presented the first of his famous **triviality results**, which have since been deepened and extended in a steady stream of subsequent work. Here I only summarize the first and simplest of the results. It should be kept in mind, however, that many attempts to circumvent them within the general truth-conditional framework have been tried and shown to fail.

Suppose, then, that the probabilistic (RT) holds and conditionals denote propositions in the usual sense – i.e., sets of worlds, or in the present framework, characteristic functions thereof. Then, since the denotations of conditionals are propositions, one can ask how their probabilities should be affected by conditionalization on other propositions: Technically, since $V(\psi \rightarrow \varphi)$ denotes a random variable, we should be able to derive its conditional expectation given that some arbitrary sentence is true.

Lewis (1976) was interested in the case that the conditioning sentence is φ , the consequent of the conditional, or its negation. For those cases, Lewis made a pair of assumptions which seem rather plausible: First, given φ , the probability of the conditional $\psi \rightarrow \varphi$ should be 1; and second, given $\bar{\varphi}$, it should be 0. Thus the equalities in (6) should hold.⁴

guments); but it is fair to say that, especially in the philosophical tradition, such proposals tend to be driven by frustration with technical obstacles (more on this in the next subsection), rather than pre-theoretical judgments. Empirically, the probabilistic interpretation of (RT) has strong and growing support (Evans and Over, 2004; Oaksford and Chater, 1994, 2003, 2007).

⁴ Lewis argued for the plausibility of (6) by invoking the **Import-Export Principle**, which in its probabilistic version requires that $P(\psi \rightarrow \varphi|\chi)$ be equivalent to $P(\varphi|\psi\chi)$. I avoid this move here because this principle is not universally accepted (see, for instance, Adams, 1975; Kaufmann, 2009).

- (6) a. $P(\psi \rightarrow \varphi|\varphi) = 1$
- b. $P(\psi \rightarrow \varphi|\bar{\varphi}) = 0$

In the present framework, this amounts to the equalities in (7).

- (7) a. $E[V(\psi \rightarrow \varphi)|V(\varphi) = 1] = 1$
- b. $E[V(\psi \rightarrow \varphi)|V(\varphi) = 0] = 0$

Now, since we are assuming that all sentences denote variables with range $\{0, 1\}$, this should hold for conditionals as well. But then (7) implies that with probability 1, the conditional is equivalent to its consequent! This consequence is grossly counterintuitive. As Lewis already pointed out, it implies that the conditional consequent is probabilistically independent of the antecedent, when in fact conditionals are typically used to convey that they are not.

The literature on triviality that ensued after the publication of Lewis’s seminal argument is vast and varied, and this is not the place to do justice to its depth and breadth. See Hájek and Hall (1994); Hájek (1994, 2012); Edgington (1995); Bennett (2003) and references therein for overviews, and Kaufmann (2005) for more discussion in the present framework. In the next section, I turn away from this literature of mostly negative results to an approach which evades triviality.

4 Stalnaker Bernoulli models

The challenge posed by the triviality results has proven to be formidable; but it is so only under certain assumptions which Lewis and many subsequent authors either took for granted or found too valuable to give up. Specifically, the assumptions are that (i) the denotations of conditionals are propositions in the usual sense – here, sets of possible worlds (more precisely, their characteristic functions); and (ii) the values of conditionals at individual worlds are fixed and do not depend on the probability distribution. Giving up these assumptions opens up an elegant way around the triviality results. This was first pointed out by van Fraassen (1976). Jeffrey (1991) arrived at a similar approach (in spirit, if not in detail) from a different angle. The connection was made explicit by Stalnaker and Jeffrey (1994), which in turn inspired subsequent work including Kaufmann (2005, 2009).

4.1 Basic idea

The main innovation is a move from possible worlds to **sequences** of possible worlds as the entities at which sentences receive their truth values and which constitute the points in the probabilistic sample space. The idea was given a compelling intuitive motivation by Stalnaker and Jeffrey (1994), with reference to Stalnaker’s (1968) similarity-based interpretation of conditionals.

Stalnaker’s proposal was that the truth value of ‘*if A, then C*’ at a possible world w is the truth value of its consequent C at a world “at which A is true and which otherwise differs minimally from the actual world.” What exactly such a

theory predicts then depends on the notion of minimal difference. For instance, Stalnaker maintained that each world is maximally similar to itself; thus if A is true at w , then no alternative world enters the evaluation because w is the most similar A -world to itself. Another assumption favored by Stalnaker was that for any world w and proposition X (including the contradiction), there is a unique X -world that is most similar to w . This assumption allowed Stalnaker to ensure that the conditional had certain logical properties which he considered desirable, especially Conditional Excluded Middle, i.e., that for any pair A, C and world w , one of ‘*if A, then C*’ and ‘*if A, then not-C*’ is true at w .

Now, the assumption that there is a unique most similar world for each proposition has always been controversial.⁵ If we give it up, allowing for multiple maximally similar worlds, but still insist on evaluating the conditional consequent relative to a single antecedent-world, then the choice of world becomes essentially non-deterministic. Van Fraassen (1976) proposed a simple formal model of such a process: If we are at an antecedent-world, we are done choosing and evaluate the conditional by evaluating its consequent. Otherwise, we continue to choose worlds in a sequence of random trials (with replacement), where at each trial the probabilities that particular worlds will be chosen are determined by the original probability distribution. Thus the probabilities at each trial are independent and identically distributed. Hence van Fraassen’s term “Stalnaker Bernoulli model,” suggestive of the special utility of this kind of model in linking the intuitions behind Stalnaker’s possible-worlds semantics for conditionals to a standard probability-theoretic framework.

4.2 Implementation

Following van Fraassen (1976) I start with a probability model and define a **product space** consisting of sets of denumerable sequences of possible worlds, along with a probability measure that is derived from the original one. As before, sentences of the language are mapped to (characteristic functions of) sets in the probability space; now these are sets of world sequences, rather than sets of worlds. The interpretation function is derived in the following way: Sentences in \mathcal{L}_A^0 are evaluated at world sequences in terms of the first world. Conditionals $\psi \rightarrow \varphi$ are evaluated at a sequence by eliminating the (possibly empty) initial sub-sequence of worlds at which the antecedent is false, then evaluating the consequent at the remaining “tail” of the sequence. At sequences at which the antecedent is false throughout, the value is undefined. The details are given in the following definition.⁶ The details are given in the following definition.

⁵ See Lewis (1973); Stalnaker (1981) for some relevant arguments.

⁶ In van Fraassen’s original version, a conditional is true, rather than undefined, at a sequence not containing any tails at which the antecedent is true. The difference is of no consequence for the cases I discuss here. In general, I find the undefinedness of the conditional probability in such cases intuitively plausible and preferable, as it squares well with widely shared intuition (in the linguistic literature, at least) that indicative conditionals with impossible antecedents give rise to presupposition

Definition 6 (Stalnaker Bernoulli model for \mathcal{L}_A). Let $\langle \Omega, \mathcal{F}, \text{Pr}, V \rangle$ be a probability model for \mathcal{L}_A^0 . The corresponding Stalnaker Bernoulli model for \mathcal{L}_A is the tuple $\langle \Omega^*, \mathcal{F}^*, \text{Pr}^*, V^* \rangle$ such that:

- Ω^* is the set of denumerable sequences of worlds in Ω .
For ω^* in Ω^* , I use the following notation:
 - $\omega^*[n]$ is the n -th world in ω^* (thus $\omega^*[n] \in \Omega$)
 - ω^{n*} is the tail of ω^* starting at $\omega^*[n]$ (thus $\omega^{n*} \in \Omega^*$)
- \mathcal{F}^* is the set of all products $X_1 \times \dots \times X_n \times \Omega^*$, for $n \geq 1$ and $X_i \in \mathcal{F}$.
- $\text{Pr}^*(\cdot)$ is a probability measure on \mathcal{F}^* defined as follows, for $X_i \in \mathcal{F}$:

$$\text{Pr}^*(X_1 \times \dots \times X_n \times \Omega^*) = \text{Pr}(X_1) \times \dots \times \text{Pr}(X_n)$$

- V^* maps pairs of sentences in \mathcal{L}_A and sequences in Ω^* to values in $\{0, 1\}$ as follows, for $\varphi, \psi \in \mathcal{L}_A$:

$$\begin{aligned} \text{if } \varphi \in \mathcal{L}_A^0, \text{ then } V^*(\varphi)(\omega^*) &= V(\varphi)(\omega^*[1]) \\ V^*(\bar{\varphi})(\omega^*) &= 1 - V^*(\varphi)(\omega^*) \\ V^*(\varphi\psi)(\omega^*) &= V^*(\varphi)(\omega^*) \times V^*(\psi)(\omega^*) \\ V^*(\varphi \rightarrow \psi)(\omega^*) &= V^*(\psi)(\omega^{n*}) \\ &\text{for the least } n \text{ s.t. } V^*(\varphi)(\omega^{n*}) = 1 \end{aligned}$$

I use boldfaced letters like ‘**X**, **Y**’ to refer to elements in \mathcal{F}^* . Notice that even if \mathcal{F} is the powerset of Ω , \mathcal{F}^* is not the powerset of Ω^* , but a proper subset of the latter. For instance, consider two arbitrary worlds ω_1, ω_2 and two arbitrary sequences $\omega_a^* = \langle \omega_1, \omega_2, \dots \rangle$ and $\omega_b^* = \langle \omega_2, \omega_1, \dots \rangle$. The set $\{\omega_a^*, \omega_b^*\}$ is in the powerset of Ω^* but not in \mathcal{F}^* , although it is a **subset** of sets in \mathcal{F}^* . This is not a deficiency for the purposes that I am putting this model to, since the truth conditions do not allow for the case that, say, a sentence is true at all and only the sequences in $\{\omega_a^*, \omega_b^*\}$.

As before, the probabilities assigned to sentences are the expectations of their truth values. Recall that sentences containing conditionals differ from sentences in \mathcal{L}_A^0 in that their values are not guaranteed to be defined: If a given sequence does not contain any A -worlds, then the value of any conditional with antecedent A at that sequence is undefined, and so is the value of compounds containing the conditional in question. The assignment of probabilities to sentences therefore comes with a qualification:

Definition 7 (Probabilities of sentences). Given a Stalnaker Bernoulli model for \mathcal{L}_A (see Definition 6), a function P^* maps sentences to real numbers as follows:

$$P^*(\varphi) = E^*[V^*(\varphi) | V^*(\varphi) \in \{0, 1\}]$$

failure. Moreover, it follows from the results below that the (un)definedness is fairly well-behaved, in the sense that the value of the conditional is defined with probability zero or one, according as the probability of the antecedent is non-zero or zero.

The conditioning event on the right-hand side in this definition is that the value of the sentence in question is defined. It turns out that this is not a very constraining condition, as the following outline of an argument shows (see also van Fraassen, 1976; Kaufmann, 2009).

First, from the definition of V^* it follows immediately that the values of sentences in \mathcal{L}_A^0 – i.e., sentences not containing any occurrences of the conditional connective – are defined at all sequences. Consider next a “first-order” conditional – that is, a sentence of the form $\varphi \rightarrow \psi$, where both φ and ψ are in \mathcal{L}_A^0 . There is only one reason why the value of this sentence at a sequence ω^* could be undefined, namely that its antecedent φ is false throughout ω^* (i.e., at $\omega^*[n]$ for all n). Proposition 1 establishes that whenever the antecedent has positive probability, the set of sequences with this property has probability 0. (Proofs of this and all subsequent results are given in the Appendix.)

Proposition 1 (van Fraassen, 1976). *For $X \in \mathcal{F}$, if $\Pr(X) > 0$, then $\Pr^*\left(\bigcup_{n \in \mathbb{N}} (\overline{X}^n \times X \times \Omega^*)\right) = 1$.*

As an immediate consequence, the value of the conditional is **almost surely** defined when its antecedent has positive probability. By similar reasoning, the value of the conditional is **almost surely** undefined if the antecedent has zero probability.⁷

As it stands, Proposition 1 only applies to elements of \mathcal{F} , i.e., denotations of sentences in \mathcal{L}_A^0 . It can be generalized to other sets, in particular the denotations of conditionals, but I will not pursue this argument here because the points I aim to make in this paper can be made with reference to simple sentences. Henceforth, to forestall any complexities of exposition arising from the possibility of undefinedness, I will implicitly limit the discussion to sentences, all of whose constituents have positive probability. With this assumption, their values are guaranteed to be defined almost surely, thus the conditioning event on the right-hand side of the equation in Definition 7 can be ignored.

Van Fraassen (1976) proved the following “Fraction Lemma,” which turns out to be useful in calculating the probabilities of both simple and compounded conditionals.

Lemma 1 (van Fraassen, 1976). *If $\Pr(X) > 0$, then $\sum_{n \in \mathbb{N}} \Pr(\overline{X})^n = 1/\Pr(X)$.*

As a straightforward consequence of the foregoing results, it is easy to ascertain that the probabilities of first-order conditionals are the corresponding conditional probabilities.

Theorem 1 (van Fraassen, 1976). *For $A, C \in \mathcal{L}_A^0$, if $\Pr(A) > 0$, then $\Pr^*(A \rightarrow C) = \Pr(C|A)$.*

⁷ In probability theory, an event happens “almost surely” if its probability is 1. This notion should not be confused with logical necessity.

In addition, it is easy to see that for all $\varphi \in \mathcal{L}_A^0$, $P^*(\varphi) = P(\varphi)$. Specifically, this is true for AC and A . Thus as an immediate corollary of Theorem 1, we have that

$$(8) \quad P^*(A \rightarrow C) = P^*(C|A)$$

This is why the interpretation in a Stalnaker Bernoulli model is not subject to the Lewis-style triviality results: The denotation of the conditional is **both** a proposition (specifically, the characteristic function of a set of sequences) **and** the corresponding conditional expectation.

Theorem 1 illustrates the simplest case of a correspondence between the probabilities assigned to conditional sentences in a Stalnaker Bernoulli model and the probabilities that their \mathcal{L}_A^0 -constituents receive in the original probability model. Similar calculations, albeit of increasing complexity, can be carried out for sentences of arbitrary complexity. This was hinted at by Stalnaker and Jeffrey (1994) and worked out for cases of conditional antecedents and consequents by Kaufmann (2009). The formulas in (9a) through (9d) serve as illustration; for more details and proofs, the reader is invited to consult Kaufmann (2009).

$$(9) \quad \text{For all } A, B, C, D \text{ in } \mathcal{L}_A^0:$$

- a. If $P(A) > 0$ and $P(C) > 0$, then
$$P^*((A \rightarrow B) \wedge (C \rightarrow D)) = \frac{P(ABCD) + P(D|C) + P(AB\bar{C}) + P(B|A)P(CD\bar{A})}{P(A \vee C)}$$
- b. If $P(A) > 0$, $P(C) > 0$, and $P(B|A) > 0$, then
$$P^*((A \rightarrow B) \rightarrow (C \rightarrow D)) = \frac{P(ABCD) + P(D|C)P(AB\bar{C}) + P(B|A)P(\bar{A}CD)}{P(A \vee C)P(B|A)}$$
- c. If $P(B) > 0$ and $P(C) > 0$, then
$$P^*(B \rightarrow (C \rightarrow D)) = P(CD|B) + P(D|C)P(\bar{C}|B)$$
- d. If $P(A) > 0$ and $P(B|A) > 0$, then
$$P^*((A \rightarrow B) \rightarrow D) = P(D|AB)P(A) + P(D\bar{A})$$

More complex compounds also receive probabilities under this approach, but they are beyond the scope of this paper because it is not clear what the empirical situation is in those cases.

4.3 Back to probability models

The preceding subsection showed how to start from an ordinary probability model and obtain values and probabilities for sentences in the derived Stalnaker Bernoulli model, defined by V^* and Pr^* . Now Stalnaker and Jeffrey (1994) (and, following them, Kaufmann, 2009) did not stop there, for their ultimate goal was to interpret sentences in \mathcal{L}_A^0 **in the original probability model** – i.e., to extend the value assignment function V from \mathcal{L}_A^0 to \mathcal{L}_A .

Intuitively, the idea is to let $V(\varphi)(\omega)$ be the conditional expectation of $V^*(\varphi)$, given the set of sequences in Ω^* whose first world is ω^* – that is, the set

$\{\omega^* \in \Omega^* | \omega^*[1] = \omega\}$. But this set may well have zero probability, since its probability is defined as $\Pr(\{\omega\})$. Stalnaker and Jeffrey’s workaround is to define the expectation relative to the ω^* -subspace of Ω^* , obtained by effectively “ignoring” the first position of the sequences.

Definition 8. *Given a probability model $\mathcal{M} = \langle \Omega, \mathcal{F}, \Pr, V \rangle$ for \mathcal{L}_A^0 , the interpretation function is extended to \mathcal{L}_A as follows: For all $\varphi \in \mathcal{L}_A$ and $\omega \in \Omega$,*

$$V(\varphi)(\omega) = \Pr^* (\{\omega^{1*} | \omega^*[0] = \omega \text{ and } V^*(\varphi)(\omega^*) = 1\})$$

where E^*, V^* are defined relative to the Stalnaker Bernoulli model based on \mathcal{M} .

It is then possible to calculate, for sentences in \mathcal{L}_A , the values they receive under V . The simple case of first-order conditionals is given in (10).

$$(10) \quad V(A \rightarrow C)(\omega) = \begin{cases} 1 & \text{if } V(A)(\omega) = V(C)(\omega) = 1 \\ 0 & \text{if } V(A)(\omega) = 1, V(C)(\omega) = 0 \\ P(C|A) & \text{if } V(A) = 0 \end{cases}$$

The first two cases on the right-hand-side of (10) are straightforward: The value of $V^*(A \rightarrow C)$ is true at either all or none of the sequences starting with an A -world, depending on the value of C at the first world. The third case is obtained by observing that the set of sequences starting with a \bar{A} -world at which the conditional is true is the union of all sets of sequences consisting of n \bar{A} -worlds followed by an AC -world, for $n \in \mathbb{N}$. By an argument along the same lines as the proof of Lemma 1 above, the measure of this set under \Pr^* is just the conditional probability of C , given A :

$$(11) \quad \sum_{n \in \mathbb{N}} (P(\bar{A})^n \times P(AC)) = P(AC)/P(A) \text{ by Lemma 1}$$

Finally, we can show that the expectation of the values in (10) is the conditional probability of the consequent, given the antecedent, as desired:

$$(12) \quad \begin{aligned} P(A \rightarrow C) &= E[V(A \rightarrow C)] \\ &= 1 \times P(AC) + 0 \times \Pr(A\bar{C}) + P(C|A) \times \Pr(\bar{A}) \\ &= P(C|A)[P(A) + P(\bar{A})] = P(C|A) \end{aligned}$$

More complex compounds involving conditionals have more complicated definitions of their value assignment. Kaufmann (2009) gives the details for the cases discussed in (9a) through (9d) above. The reader is referred to the paper for details.

4.4 Interim summary

The foregoing outlines a strategy for calculating values and probabilities for sentences in \mathcal{L}_A relative to a probability model \mathcal{M} , pictured schematically in Figure 1. The extension from \mathcal{L}_A^0 to \mathcal{L}_A crucially requires the construction of

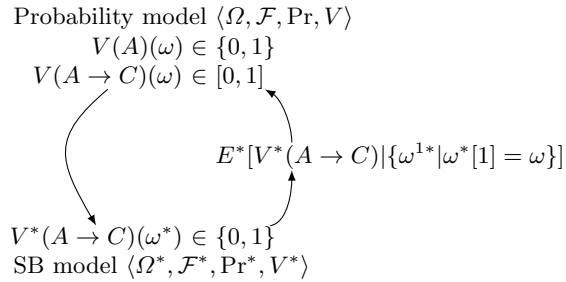


Fig. 1. Value assignment for conditionals via the derived SB model

the Stalnaker Bernoulli model \mathcal{M}^* , which is then used in the interpretation of sentences involving conditionals.

Now, it is clear that this clever strategy is also somewhat roundabout. All sentences receive values and probabilities in both models; moreover, due to the way one is defined in terms of the other, the probabilities actually coincide: $P(\varphi) \equiv P^*(\varphi)$ for all φ . This naturally raises the question why both models are required.

Little reflection is needed to see that once we raise this issue, it comes down to the question whether we still need the simpler probability model. For it is clear that the Stalnaker Bernoulli model is indispensable: It is crucially involved in the derivation of values and probabilities for conditionals and more complex sentences containing conditional constituents.⁸ But then, what is the utility of having the traditional, world-based model?

I will not venture to offer a conclusive answer to this question at this point; surely one factor in favor of the conventional possible-worlds model is its familiarity and relatively straightforward and well-understood interface with general topics in epistemology and metaphysics. So the question is to what extent these connections could be recreated in a framework that was primarily based on the Stalnaker Bernoulli approach.

Aside from such big-picture considerations, it is worth stressing that a general shift to the Stalnaker Bernoulli framework would bring additional functionality beyond the extension of the interpretation to \mathcal{L}_A . To wit, adopting the approach of assigning intermediate values to conditionals at worlds at which their antecedents are false, we face the problem that there is no general rule of conditionalization involving conditionals – either for conditioning their denotations on other events, or for conditioning other events on them. This problem

⁸ As a historical side note, it is worth pointing out that some of the functionality delivered here by the Stalnaker Bernoulli model can also be achieved in a simpler model. This was shown by Jeffrey (1991), who developed the random-variable approach with intermediate truth values without relying on van Fraassen’s construction. But that approach has its limits, for instance when it comes to conditionals with conditional antecedents, and can be seen as superseded by the Stalnaker Bernoulli approach.

does not arise in the Stalnaker Bernoulli approach, where the probability of a conditional is the probability of a proposition (in addition to being a conditional probability).

To get such a project off the ground, however, much work has to be done. For one thing, the Stalnaker Bernoulli model would need some intuitive interpretation. Moreover, we would need plausible representations of such elementary notions as belief change by conditionalization.

Regarding the first of these desiderata – an intuitive interpretation of Stalnaker Bernoulli models – we can derive some inspiration from the suggestions of Stalnaker and Jeffrey, cited above. A set \mathbf{X} of world sequences represents two kinds of information: “factual” beliefs are encoded in the set of “first worlds” $\{\omega^*[1] \mid \omega^* \in \mathbf{X}\}$, which is where all sentences in \mathcal{L}_A^0 receive their truth values. The other kind of information consists in “conditional” beliefs, encoded in the sequences $\{\omega^{2*} \mid \omega^* \in \mathbf{X}\}$. Each sequence represents a possible outcome of a countable sequence of random choices of a world (with replacement), modeling a non-deterministic variant of the interpretation of conditionals in terms of a Stalnaker selection function.⁹

With this in mind, I will set the first issue aside and spend the rest of this paper focusing on the second issue, the definition of belief update by conditionalization in a Stalnaker Bernoulli model.

5 Conditionalization

What is the problem with conditionalization in Stalnaker Bernoulli models? Figure 2 gives a general overview. Suppose we start out, as we did in this paper, with \mathcal{M}_1 , a conventional probabilistic possible-worlds model. In it, the non-conditional sentences in \mathcal{L}_A^0 receive truth values at worlds, but in order to extend the value assignment to conditionals in accordance with the probabilistic interpretation of (RT), we take a detour via the derived Stalnaker Bernoulli model \mathcal{M}_1^* (as shown in Figure 1 above). This allows us to extend V_1 to conditionals.

Now, suppose we want to update the belief model by conditioning on a new piece of information, say, that some sentence φ is true. If $\varphi \in \mathcal{L}_A^0$, this is not a problem: As usual in Bayesian update, we conditionalize by shifting the probability mass to the set of worlds at which φ is true, then renormalize the measure so as to ensure that we have once again a probability distribution, thus obtaining Pr_2 .

Now, a number of things are noteworthy about this update operation. First, along with the shift from Pr_1 to Pr_2 , the value assignment for conditionals must also change in order to ensure that the expectation of the new measure is the conditional probability. Thus we obtain a valuation function V_2 which agrees with V_1 on all sentences in \mathcal{L}_A^0 , but may differ in the values it assigns to conditionals and sentences containing them. The second noteworthy point about the update

⁹ In a related sense, one may also think of a given set of sequences as representing all paths following an introspective (i.e., transitive and euclidean) doxastic accessibility relation. I leave the further exploration of this connection for future work.

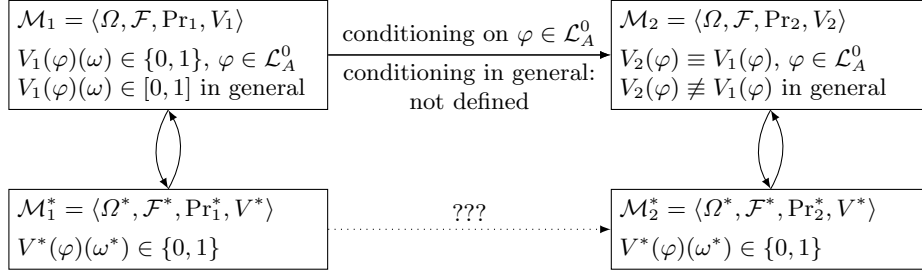


Fig. 2. Belief update in probability models with SB interpretations for conditionals

is that we have no systematic way of deriving this new assignment function V_2 from V_1 . After the update, the new values for conditionals have to be filled in by once again taking the detour via the derived Stalnaker Bernoulli model, now \mathcal{M}_2^* . Thirdly, while the update with a sentence in \mathcal{L}_A^0 just discussed merely presents a minor inconvenience (in calling for a recalculation of the values of sentences not in \mathcal{L}_A^0 via the derived Stalnaker Bernoulli model), matters are worse if the incoming information is conditional, i.e., the information that some sentence not in \mathcal{L}_A^0 is true. Whenever such a sentence takes values in between 0 and 1, we do not have a way to conditionalize on it.

Notice, though, that these problems would not arise if we were using Stalnaker Bernoulli models throughout. In \mathcal{M}_1^* and \mathcal{M}_2^* , all sentences receive values in $\{0, 1\}$ almost surely (with the above caveats about the possibility of undefinedness and its intuitive justification). Moreover, notice that the two Stalnaker Bernoulli models share the same valuation function. For recall that $V^*(\varphi)(\omega^*)$ only depends on the structure of ω^* , not on the probability distribution. And since conditionals denote propositions in these models (i.e., elements of \mathcal{F}^*), conditionalization involving those denotations can proceed in the familiar fashion. I take all of these facts to be compelling arguments in favor of the exclusive use of Stalnaker Bernoulli models.

5.1 Shallow conditioning

How exactly should belief update be carried out in Stalnaker Bernoulli models? To see why this is an issue, consider what would seem to be the most straightforward way to define conditionalization on a sentence φ in \mathcal{M}_1^* . Presumably, similarly to the analogous procedure on \mathcal{M}_1 , this would involve shifting the probability mass onto sequences at which $V^*(\varphi)$ evaluates to 1, then renormalizing the measure to ensure that the result is again a probability distribution.¹⁰

¹⁰ An alternative way of achieving the same result would be to model belief update in terms of “truncation” of world sequences along the lines of the interpretation of conditionals, chopping off initial sub-sequences until the remaining tail verifies φ . I will not go into the details of this operation here; it corresponds to the operation of shallow conditioning discussed in this subsection, for the same reason that the probabilities of conditionals equal the corresponding conditional probabilities.

Now, the problem with this approach is that the resulting probability distribution is not \Pr_2^* , hence the resulting model is not \mathcal{M}_2^* . It is easy to see why this is the case. For concreteness, let us assume that the new information is that some sentence A in \mathcal{L}_A^0 is true. By the definition of the Stalnaker Bernoulli model, $V^*(A)(\omega^*) = 1$ whenever $V(A)(\omega^*[1]) = 1$ – in words, if **the first world** in ω^* is an A -world. This includes all sequences which begin with an A -world, regardless of what happens later on in them.

But this is not the set of sequences over which the probability is distributed in \mathcal{M}_2^* . For recall that \mathcal{M}_2^* was obtained from \mathcal{M}_1 by conditionalization on the information that A is true. As a result of this operation, in \mathcal{M}_2 the entire probability mass is assigned to worlds at which A is true. Hence, in \mathcal{M}_2^* , the probability mass is concentrated on sequences which consist **entirely** of A -worlds (i.e., sequences ω^* such that $V^*(A)(\omega^{n*}) = 1$ for all n).

Thus in order to fill in the missing step from \mathcal{M}_1^* to \mathcal{M}_2^* in Figure 2, we would have to conditionalize on the set of sequences in (13b), rather than (13a).

$$(13) \quad \begin{array}{l} \text{a. } \{\omega^* \in \Omega^* \mid V^*(\varphi)(\omega^*) = 1\} \\ \text{b. } \{\omega^* \in \Omega^* \mid V^*(\varphi)(\omega^{n*}) = 1 \text{ for all } n \geq 1\} \end{array}$$

However, the set in (13b) has zero probability whenever the probability of φ is less than 1. Indeed, for any $X \in \mathcal{F}$, the set X^* of sequences consisting entirely of X -worlds has probability 1 or 0, according as the probability of X is 1 or less than 1:

$$(14) \quad \Pr^*(X^*) = \lim_{n \rightarrow \infty} \Pr(X)^n = \begin{cases} 1 & \text{if } \Pr(X) = 1 \\ 0 & \text{otherwise} \end{cases}$$

Clearly a different definition is needed to work around this problem.

5.2 Deep conditioning

It is not uncommon in discussions of probability to see “Bayes’s Rule” as a definition of conditional probability:

$$(BR) \quad \Pr(X|Y) = \frac{\Pr(X \cap Y)}{\Pr(Y)} \text{ when } \Pr(Y) > 0$$

Taken as a definition, this suggests that the conditional probability is generally undefined when the probability of the conditioning event is 0. But this view has many problematic consequences, not the least of them being that it predicts that the conditional probability is undefined in cases in which we in fact have clear intuitions that it exists and what it should be.¹¹ This point has been discussed

¹¹ As a simple example, consider the task of choosing a point (x, y) at random from a plane. Fix some point (x^*, y^*) and consider the conditional probability that $y > y^*$, given $x = x^*$ (intuitively, the conditional probability that the randomly chosen point will lie above (x^*, y^*) , given that it lies on the vertical line through (x^*, y^*)). We have clear intuitions as to what this conditional probability is and how it depends

in the philosophical literature (see for instance Stalnaker and Jeffrey, 1994; Jeffrey, 2004; Hájek, 2003, 2011, and references therein), but the misconception that (BR) is the definition of conditional probability is apparently hard to root out.¹²

My own take on conditional probability follows the lead of Stalnaker and Jeffrey (1994) and Jeffrey (2004): (BR) is not a definition, but a restatement of the “Product Rule” (PR):

$$(PR) \quad \Pr(X|Y) \times \Pr(Y) = \Pr(X \cap Y)$$

Most crucially, (PR) should not be mistaken for a definition either, but as an axiom regulating the relationship between unconditional and conditional probabilities **when both are defined**.

Following this general strategy, in Definition 9 I give one definition of a two-place function $\Pr^*(\cdot|\cdot)$, then I proceed to show that it is in fact properly called a conditional probability.

Definition 9 (Stalnaker Bernoulli model with conditional probability). *An Stalnaker Bernoulli model with conditional probability is a tuple $\langle \Omega^*, \mathcal{F}^*, \Pr^*(\cdot), \Pr^*(\cdot|\cdot), V^* \rangle$, where $\langle \Omega^*, \mathcal{F}^*, \Pr^*(\cdot), V^* \rangle$ is a Stalnaker Bernoulli model (see Definition 6) and $\Pr^*(\cdot|\cdot)$ is a partial function mapping pairs of propositions in \mathcal{F}^* to real numbers as follows, for all \mathbf{X}, \mathbf{Y} in \mathcal{F}^* :*

$$\Pr^*(\mathbf{X}|\mathbf{Y}) = \lim_{n \rightarrow \infty} \frac{\Pr^*(X_1 \cap Y_1 \times \dots \times X_n \cap Y_n \times \Omega^*)}{\Pr^*(Y_1 \times \dots \times Y_n \times \Omega^*)}$$

This definition opens up the possibility that $\Pr^*(\mathbf{X}|\mathbf{Y})$ is defined while the ratio $\Pr^*(\mathbf{X} \cap \mathbf{Y})/\Pr^*(\mathbf{Y})$ is not. For example, let $\mathbf{X} = \mathbf{Y} = Z^*$ for some $Z \in \mathcal{F}$ with $0 < \Pr(Z) < 1$, and note that in this case $\Pr^*(Z^*) = \lim_{n \rightarrow \infty} \Pr(Z)^n = 0$. Thus for instance, the quotient $\Pr^*(Z^*)/\Pr^*(Z^*)$ is undefined, hence Bayes’s Rule is silent on the conditional probability of Z^* given Z^* . However, under Definition 9 the value of $\Pr^*(Z^*|Z^*)$ is defined (proofs are given in the appendix):

Proposition 2. *If $\Pr(Z) > 0$, then $\Pr^*(Z^*|Z^*) = 1$.*

Now, while $\Pr^*(\cdot|\cdot)$ may be defined when the quotient in (BR) is not, the next two results show that it is “well-behaved” with respect to the Product rule (PR).

Proposition 3. *If $\Pr^*(\mathbf{X}|\mathbf{Y})$ is defined, then $\Pr^*(\mathbf{X}|\mathbf{Y}) \times \Pr^*(\mathbf{Y}) = \Pr^*(\mathbf{X} \cap \mathbf{Y})$.*

on the location of the cutoff point; but the probability that the randomly chosen point lies on the line is 0.

¹² Notice, incidentally, that the view on conditional probability just endorsed is not at odds with the remarks on the undefinedness of the values of conditionals at world sequences throughout which the antecedent is false (see Footnote 6 above). For one thing, technically the undefinedness discussed there does not enter the picture because some conditional probability is undefined. But that aside, I emphasize that I do not mean to claim that conditional probabilities given zero-probability events are **always** defined, but only that they can be.

As mentioned above, Stalnaker and Jeffrey (1994) likewise prefer to impose the product rule as an axiom, rather than using it as a definition. They also impose an additional condition which in the present framework falls out from the definition of conditional probability whenever it is defined:

Proposition 4. *If $\Pr^*(\mathbf{X}|\mathbf{Y})$ is defined and $\mathbf{Y} \subseteq \mathbf{X}$, then $\Pr^*(\mathbf{X}|\mathbf{Y}) = 1$.*

I conclude that $\Pr^*(\cdot|\cdot)$ can properly be called a “conditional probability.”¹³ At the same time, there are cases in which (BR) cannot be applied, yet $\Pr^*(\cdot|\cdot)$ is defined. This possibility is important in modeling belief update in Stalnaker Bernoulli models.

The remaining results in this section show that conditionalization according to Definition 9 is well-behaved. The first generalizes Proposition 1 above to probabilities conditionalized on a sequence Z^* .

Proposition 5. *If $\Pr(X \cap Z) > 0$, then $\Pr^*\left(\bigcup_{n \in \mathbb{N}} (\bar{X}^n \times X \times \Omega^*) \middle| Z^*\right) = 1$.*

Notice that Proposition 1 is a corollary of Proposition 5, substituting Ω for Z . Lemma 1 (van Fraassen’s “Fraction Lemma”) can likewise be generalized to conditional probabilities.

Lemma 2 (Conditional Fraction Lemma). *If $\Pr(X \cap Z) > 0$, then $\sum_{n \in \mathbb{N}} \Pr(\bar{X}|Z)^n = 1/\Pr(X|Z)$.*

Lemma 2 allows us to determine the conditional probability assigned to the denotation of a conditional given a sequence Z^* .

6 Some consequences for conditionals

To sum up the preceding subsections, the Stalnaker Bernoulli framework offers two natural ways to interpret conditionals: by the rule in Definition 6, and by deep conditioning. The two are distinguished by the characterization of the conditioning event, i.e., the way in which the incoming information is used in singling out the set of sequences on which to concentrate the probability mass. For a given sentence φ , the two options are repeated in (17). I argued above that deep conditioning is what is required to model belief update by conditionalization.

$$(17) \quad \begin{array}{ll} \text{a.} & \{\omega^* \in \Omega^* | V^*(\varphi)(\omega^*) = 1\} & [\textit{shallow}] \\ \text{b.} & \{\omega^* \in \Omega^* | V^*(\varphi)(\omega^{n*}) = 1 \text{ for all } n\} & [\textit{deep}] \end{array}$$

In addition, I believe it is likely that **both** update operations are required to model the interpretation of conditionals. To see this, consider first what the difference comes down to in terms of the intuitive interpretation of the two versions of update.

¹³ However, $\Pr^*(\mathbf{X}|\mathbf{Y})$ is itself not always defined: It is undefined if either $\Pr(Y_i) = 0$ for any i , or the function does not converge as n approaches infinity.

Recall first the intuitive interpretation of a Stalnaker Bernoulli model as a representation of epistemic states. The idea was that a world sequence represents factual information in the first world, and conditional information in the tail following the first world. As already mentioned in Footnote 9 above, from the modal-logic perspective on the representation of knowledge and belief, one can alternatively think of the sequences as representing all possible paths along an introspective doxastic accessibility relation. This intuitive interpretation of world sequences corresponds well with the proposal to model belief update by conditionalization on the set of sequences throughout which the incoming sentence is true (as opposed to the set of sequences at which the sentence evaluates to 1).

With all this in mind, I believe that we can discern in the distinction between shallow and deep conditioning an intuitively real difference in ways to “hypothetically add,” in Ramsey’s words, a sentence φ to one’s probabilistic “stock of knowledge”: Shallow update consists in assuming that φ is **true**, whereas deep update consists in assuming that φ is **learned**.

The distinction between these two modes of update underlies certain problematic examples which have been discussed widely in the literature on conditionals. Among them are simple conditionals like (18a), attributed to Richmond Thomason by van Fraassen (1980), and (18b) from Lewis (1986).

- (18) a. If my wife deceives me, I won’t believe it.
 b. If Reagan works for the KGB, I’ll never believe it.

In both cases, the conditional antecedent is intuitively not correctly paraphrased by any variant of the locution ‘*If I know / learn that . . .*’, for otherwise it would not make sense together with the consequent. Yet these examples are perfectly well-formed and interpretable. What the examples show is that it is possible for speakers to suppose that something is the case unbeknownst to them. The present framework opens up a way to model the difference in a probabilistic setting.

Another difference between the two modes of update concerns the interpretation of right-nested conditionals, i.e., sentences of the form $B \rightarrow (C \rightarrow D)$. Here we see a tangible consequence of the fact that the update with the antecedent B determines the context for the interpretation of the consequent $D \rightarrow D$. Theorem 2 states the general result for deep conditioning with sentences of this form.

Theorem 2. *If $P(BC) > 0$, then $P^*(C \rightarrow D|B^*) = P(D|BC)$.*

Theorem 1 above is a special case of Theorem 2, again obtained by substituting the tautology for B .

What is notable about this result is that it resembles the **Conditional Import-Export Principle**, i.e., the rule that the conditional probability of $\varphi \rightarrow \psi$ given χ should equal the conditional probability of ψ given $\varphi\chi$. Theorem 2 shows that the Import-Export Principle holds for right-nested conditionals under deep conditioning on the antecedent. On the other hand, it does not hold

under the standard Stalnaker Bernoulli interpretation. Recall from (9c) above that the expectation of the values assigned to such right-nested conditionals is quite different:

$$(9c) \quad \text{If } P(B) > 0 \text{ and } P(C) > 0, \text{ then} \\ P^*(B \rightarrow (C \rightarrow D)) = P(CD|B) + P(D|C)P(\bar{C}|B)$$

There is much room for further explorations into the ramifications of this distinction, for right-nested conditionals as well as for the others listed in (9b) and (9d) above. These investigations are left to future work.

7 Conclusions

I have argued that it makes good sense to investigate the utility of Stalnaker Bernoulli models as a formal tool for the representation and analysis of belief states and their dynamics. The main arguments in favor of such a move draw on the viability of an account long these lines that is immune to Lewisian triviality results. But I also showed that once we start to investigate the matter seriously, a number of additional subtleties and advantages come into view which merit further study. The difference between shallow and deep conditioning and its potential applications in the analysis of counterexamples to the Import-Export Principle were one such example. Additionally, the framework promises to offer a straightforward account of belief updates involving conditionals, whose theoretical implementation and empirical verification remain to be carried out. And all of this, in a sense, leads up to the big question: whether it is time to do away with the time-honored but simplistic possible-worlds models, if only to get the meaning of conditionals right.

Appendix: Proofs

Proposition 1. *For $X \in \mathcal{F}$, if $\Pr(X) > 0$, then $\Pr^*\left(\bigcup_{n \in \mathbb{N}} (\bar{X}^n \times X \times \Omega^*)\right) = 1$.*

Proof. Notice that $\bigcup_{n \in \mathbb{N}} (\bar{X}^n \times X \times \Omega^*)$ is the set of all sequences containing at least one X -world, thus its complement is \bar{X}^* . Now $\Pr^*\left(\bigcup_{n \in \mathbb{N}} (\bar{X}^n \times X \times \Omega^*)\right) = 1 - \Pr^*(\bar{X}^*)$
 $= 1 - \lim_{n \rightarrow \infty} \Pr^*(\bar{X}^n \times \Omega^*) = 1 - \lim_{n \rightarrow \infty} \Pr(\bar{X})^n = 1$ since $\Pr(\bar{X}) < 1$. \square

Lemma 1. *If $\Pr(X) > 0$, then $\sum_{n \in \mathbb{N}} \Pr(\bar{X})^n = 1/\Pr(X)$.*

Proof. $\sum_{n \in \mathbb{N}} \Pr(\bar{X})^n \times \Pr(X) = \sum_{n \in \mathbb{N}} \left(\Pr(\bar{X})^n \times \Pr(X)\right)$
 $= \sum_{n \in \mathbb{N}} \Pr^*(\bar{X}^n \times X \times \Omega^*) = \Pr^*\left(\bigcup_{n \in \mathbb{N}} (\bar{X}^n \times X \times \Omega^*)\right) = 1$ by Prop. 1 \square

Theorem 1. For $A, C \in \mathcal{L}_A^0$, if $P(A) > 0$, then $P^*(A \rightarrow C) = P(C|A)$.

Proof. By Definition 6, the set of sequences ω^* such that $V^*(A \rightarrow C)(\omega^*) = 1$ is the union $\bigcup_{n \in \mathbb{N}} (\{\omega \in \Omega | V(A)(\omega) = 0\}^n \times \{\omega \in \Omega | V(AC)(\omega) = 1\} \times \Omega^*)$. Since the sets for different values of n are mutually disjoint, the probability of the union is the sum of the probabilities for all n . Now, for all $X, Y \in \mathcal{F}$,

$$\begin{aligned} \Pr^* \left(\bigcup_{n \in \mathbb{N}} \left(\bar{Y}^n \times (X \cap Y) \times \Omega^* \right) \right) &= \sum_{n \in \mathbb{N}} \Pr^* \left(\bar{Y}^n \times (X \cap Y) \times \Omega^* \right) \\ &= \sum_{n \in \mathbb{N}} \left(\Pr(\bar{Y})^n \times \Pr(X \cap Y) \right) = \sum_{n \in \mathbb{N}} \Pr(\bar{Y})^n \times \Pr(X \cap Y) \\ &= \Pr(X \cap Y) / \Pr(Y) \text{ by Lemma 1.} \end{aligned}$$

In particular, let X, Y be the set of worlds in Ω at which $V(C)$ and $V(A)$ are true, respectively. \square

Proposition 2. If $\Pr(Z) > 0$, then $\Pr^*(Z^*|Z^*) = 1$.

Proof. $\Pr^*(Z^*|Z^*) = \lim_{n \rightarrow \infty} \Pr^*(Z^n \times \Omega^* | Z^n \times \Omega^*) = 1$. \square

Proposition 3. If $\Pr^*(\mathbf{X}|\mathbf{Y})$ is defined, then $\Pr^*(\mathbf{X}|\mathbf{Y}) \times \Pr^*(\mathbf{Y}) = \Pr^*(\mathbf{X} \cap \mathbf{Y})$.

Proof. Since $\Pr^*(\mathbf{X}|\mathbf{Y})$ is defined, $\Pr(Y_n) > 0$ for all n . Thus $\Pr^*(\mathbf{X}|\mathbf{Y}) \times \Pr^*(\mathbf{Y})$

$$\begin{aligned} &= \lim_{n \rightarrow \infty} \frac{\Pr^*(X_1 \cap Y_1 \times \dots \times X_n \cap Y_n \times \Omega^*)}{\Pr^*(Y_1 \times \dots \times Y_n \times \Omega^*)} \times \lim_{n \rightarrow \infty} \Pr^*(Y_1 \times \dots \times Y_n \times \Omega^*) \\ &= \lim_{n \rightarrow \infty} \left(\frac{\Pr^*(X_1 \cap Y_1 \times \dots \times X_n \cap Y_n \times \Omega^*)}{\Pr^*(Y_1 \times \dots \times Y_n \times \Omega^*)} \times \Pr^*(Y_1 \times \dots \times Y_n \times \Omega^*) \right) \\ &= \lim_{n \rightarrow \infty} \Pr^*(X_1 \cap Y_1 \times \dots \times X_n \cap Y_n \times \Omega^*) = \Pr^*(\mathbf{X} \cap \mathbf{Y}) \quad \square \end{aligned}$$

Proposition 4. If $\Pr^*(\mathbf{X}|\mathbf{Y})$ is defined and $\mathbf{Y} \subseteq \mathbf{X}$, then $\Pr^*(\mathbf{X}|\mathbf{Y}) = 1$.

Proof. For all $i \geq 1$, $X_i \cap Y_i = Y_i$ since $\mathbf{Y} \subseteq \mathbf{X}$, and $\Pr(Y_i) > 0$ since $\Pr^*(\mathbf{X}|\mathbf{Y})$ is defined. Thus $\Pr^*(\mathbf{X}|\mathbf{Y}) = \lim_{n \rightarrow \infty} \frac{\Pr^*(Y_1 \times \dots \times Y_n \times \Omega^*)}{\Pr^*(Y_1 \times \dots \times Y_n \times \Omega^*)} = 1$. \square

The following auxiliary result will be useful in the subsequent proofs.

Proposition 9. If $\Pr(Z) > 0$, then for $X_i \in \mathcal{F}$,

$$\Pr^*(X_1 \times \dots \times X_n \times \Omega^* | Z^*) = \Pr^*(X_1 \times \dots \times X_n \times \Omega^* | Z^n \times \Omega^*)$$

Proof. Immediate because $Z \subseteq \Omega$. \square

The significance of Proposition 9 derives from the fact that the sets of sequences at which a given sentence in \mathcal{L}_A is true can be constructed (using set operations under which \mathcal{F}^* is closed) out of sequence sets ending in Ω^* . This is obvious for the non-conditional sentences in \mathcal{L}_A^0 . For conditionals $\varphi \rightarrow \psi$ the relevant set is the union of sets of sequences consisting of n $\bar{\varphi}$ -worlds followed by a $\varphi\psi$ -world. For each n , the corresponding set ends in Ω^* and therefore can be conditioned upon Z^* as shown in Proposition 9. Since these sets for different numbers n are mutually disjoint, the probability of their union is just the sum of their individual probabilities.

Proposition 5. *If $\Pr(X \cap Z) > 0$, then $\Pr^* \left(\bigcup_{n \in \mathbb{N}} \left(\overline{X}^n \times X \times \Omega^* \right) \middle| Z^* \right) = 1$.*

Proof. Since $\Pr(X \cap Z) > 0$, $\Pr(\overline{X} \cap Z) < \Pr(Z)$. Thus

$$\begin{aligned} \Pr^* \left(\bigcup_{n \in \mathbb{N}} \left(\overline{X}^n \times X \times \Omega^* \right) \middle| Z^* \right) &= 1 - \Pr^* \left(\overline{X}^* \middle| Z^* \right) \\ &= 1 - \lim_{n \rightarrow \infty} \frac{\Pr^* \left((\overline{X} \cap Z)^n \times \Omega^* \right)}{\Pr^* \left(Z^n \times \Omega^* \right)} = 1 - \lim_{n \rightarrow \infty} \frac{\Pr \left(\overline{X} \cap Z \right)^n}{\Pr \left(Z \right)^n} \\ &= 1 - \lim_{n \rightarrow \infty} \left(\frac{\Pr \left(\overline{X} \cap Z \right)}{\Pr \left(Z \right)} \right)^n = 1 \text{ since } \Pr^* \left(\overline{X} \cap Z \right) < \Pr^* \left(Z \right) \quad \square \end{aligned}$$

Lemma 2. *If $\Pr(X \cap Z) > 0$, then $\sum_{n \in \mathbb{N}} \Pr \left(\overline{X} | Z \right)^n = 1 / \Pr(X | Z)$.*

Proof. Since $\Pr(X \cap Z) = \Pr(Z) \times \Pr(X | Z)$, both $\Pr(Z) > 0$ and $\Pr(X | Z) > 0$.

$$\begin{aligned} \sum_{n \in \mathbb{N}} \Pr \left(\overline{X} | Z \right)^n \times \Pr(X | Z) &= \sum_{n \in \mathbb{N}} \left(\Pr \left(\overline{X} | Z \right)^n \times \Pr(X | Z) \right) \\ &= \sum_{n \in \mathbb{N}} \frac{\Pr \left(\overline{X} \cap Z \right)^n \times \Pr(X \cap Z)}{\Pr(Z)^{n+1}} = \sum_{n \in \mathbb{N}} \frac{\Pr^* \left((\overline{X} \cap Z)^n \times (X \cap Z) \times \Omega^* \right)}{\Pr^* \left(Z^{n+1} \times \Omega^* \right)} \\ &= \sum_{n \in \mathbb{N}} \Pr^* \left(\overline{X}^n \times X \times \Omega^* \middle| Z^{n+1} \times \Omega^* \right) \\ &= \sum_{n \in \mathbb{N}} \Pr^* \left(\overline{X}^n \times X \times \Omega^* \middle| Z^* \right) \text{ by Proposition 9} \\ &= \Pr^* \left(\bigcup_{n \in \mathbb{N}} \left(\overline{X}^n \times X \times \Omega^* \right) \middle| Z^* \right) = 1 \text{ by Proposition 5.} \quad \square \end{aligned}$$

Theorem 2. *If $P(BC) > 0$, then $P^*(C \rightarrow D | B^*) = P(D | BC)$.*

Proof. $P^*(C \rightarrow C | B^*) = \lim_{n \rightarrow \infty} \frac{P^* \left(((C \rightarrow D) \wedge B)^n \times \Omega^* \right)}{P^* \left(B^n \times \Omega^* \right)}$

$$\begin{aligned} &= \lim_{n \rightarrow \infty} \frac{\sum_{i=0}^{n-1} P(\overline{C} B)^i \times P(CDB)}{P(B)^n} \\ &= \lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} P(\overline{C} | B)^i \times P(CD | B) \\ &= P(CD | B) / P(C | B) \text{ by Lemma 2} \\ &= P(D | BC) \quad \square \end{aligned}$$

Bibliography

- Barbara Abbott. Some remarks on indicative conditionals. In Robert B. Young, editor, *Proceedings of SALT 14*, pages 1–19, Ithaca, NY, 2004. Cornell University.
- Ernest Adams. The logic of conditionals. *Inquiry*, 8:166–197, 1965.
- Ernest Adams. *The Logic of Conditionals*. Reidel, 1975.
- Jonathan Bennett. *A Philosophical Guide to Conditionals*. Oxford University Press, 2003.
- Dorothy Edgington. On conditionals. *Mind*, 104(414):235–329, April 1995.
- Ellery Eells and Brian Skyrms, editors. *Probabilities and Conditionals: Belief Revision and Rational Decision*. Cambridge University Press, 1994.
- Jonathan St.B.T. Evans and David E. Over. *If*. Oxford University Press, Oxford, UK, 2004.
- James H. Fetzer, editor. *Probability and Causality*, volume 192 of *Studies in Epistemology, Logic, Methodology, and Philosophy of Science*. D. Reidel, 1988.
- Bas C. van Fraassen. Probabilities of conditionals. In William L. Harper, Robert Stalnaker, and Glenn Pearce, editors, *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, volume 1 of *The University of Western Ontario Series in Philosophy of Science*, pages 261–308. D. Reidel, 1976.
- Bas C. van Fraassen. Review of Brian Ellis, “Rational Belief Systems”. *Canadian Journal of Philosophy*, 10:457–511, 1980.
- A. Gibbard. Two recent theories of conditionals. In Harper, Stalnaker, and Pearce (1981), pages 211–247.
- Alan Hájek. Triviality on the cheap? In Eells and Skyrms (1994), pages 113–140.
- Alan Hájek. What conditional probability could not be. *Synthese*, 137:273–323, 2003.
- Alan Hájek. Conditional probability. In Prasanta S. Bandyopadhyay and Malcolm R. Forster, editors, *Philosophy of Statistics*, volume 7 of *Handbook of the Philosophy of Science*. Elsevier B.V., 2011. Series editors: Dov M. Gabbay, Paul Thagard, and John Woods.
- Alan Hájek. The fall of “Adams’ Thesis”? *Journal of Logic, Language and Information*, 21(2):145–161, 2012.
- Alan Hájek and Ned Hall. The hypothesis of the conditional construal of conditional probability. In Eells and Skyrms (1994), pages 75–110.
- William L. Harper, R. Stalnaker, and G. Pearce, editors. *Ifs: Conditionals, Belief, Decision, Chance, and Time*. Reidel, 1981.
- Frank Jackson. On assertion and indicative conditionals. *Philosophical Review*, 88:565–589, 1979.
- Richard C. Jeffrey. If. *Journal of Philosophy*, 61:702–703, 1964.
- Richard C. Jeffrey. Matter-of-fact conditionals. In *The Symposia Read at the Joint Session of the Aristotelian Society and the Mind Association at the*

- University of Durham*, pages 161–183. The Aristotelian Society, July 1991. Supplementary Volume 65.
- Richard C. Jeffrey. *Subjective Probability: The Real Thing*. Cambridge University Press, 2004.
- Stefan Kaufmann. Conditional predictions: A probabilistic account. *Linguistics and Philosophy*, 28(2):181–231, 2005.
- Stefan Kaufmann. Conditionals right and left: Probabilities for the whole family. *Journal of Philosophical Logic*, 38:1–53, 2009.
- David Lewis. *Counterfactuals*. Harvard University Press, 1973.
- David Lewis. Probabilities of conditionals and conditional probabilities. *Philosophical Review*, 85:297–315, 1976.
- David Lewis. Postscript to “Probabilities of conditionals and conditional probabilities”. In *Philosophical Papers*, volume 2, pages 152–156. Oxford University Press, 1986.
- Mike Oaksford and Nick Chater. A rational analysis of the selection task as optimal data selection. *Psychological Review*, 101:608–631, 1994.
- Mike Oaksford and Nick Chater. Conditional probability and the cognitive science of conditional reasoning. *Mind & Language*, 18(4):359–379, 2003.
- Mike Oaksford and Nick Chater. *Bayesian Rationality: The Probabilistic Approach to Human Reasoning*. Oxford University Press, Oxford, UK, 2007.
- F. P. Ramsey. General propositions and causality. Printed in ?, pages 145–163, 1929.
- Robert Stalnaker. A theory of conditionals. In *Studies in Logical Theory, American Philosophical Quarterly, Monograph: 2*, pages 98–112. Blackwell, 1968.
- Robert Stalnaker. Probability and conditionals. *Philosophy of Science*, 37:64–80, 1970.
- Robert Stalnaker. A defense of conditional excluded middle. In Harper et al. (1981), pages 87–104.
- Robert Stalnaker and Richard Jeffrey. Conditionals as random variables. In Eells and Skyrms (1994), pages 31–46.