Learning 'If'

05.23.2024 // World Sequence Day, UConn // Calum McNamara*

The *Bayesian* view in epistemology has two core commitments:

*University of Michigan, Ann Arbor
Email: camcnam@umich.edu

(i) Beliefs come in degrees, and rational degrees of belief—or *credences*—obey the probability axioms.

(ii) Changes in credence are governed by *conditionalization* or *Jeffrey conditionalization* (depending on the circumstances).

Jeffrey conditionalization was introduced by Richard Jeffrey (1965). He calls it *probability kinematics*.

In the case of (ii), the relevant rules are:

Conditionalization. After learning A with certainty, your credences should satisfy $q(-) = p(- \mid A)$.

In what follows, I write 'p' for your prior credence function, and 'q' for your posterior credence function.

Jeffrey Conditionalization. Let $\{A_i\}$ be a partition. Suppose $q(A_i) \neq p(A_i)$ for some A_i. Then $q(-) = \sum_i p(- \mid A_i) \cdot q(A_i)$.

I'll assume (i). What about (ii)?

(ii) has been subject to criticism. One problem with it is that conditionalization and/or Jeffrey conditionalization seems to deliver counterintuitive verdicts in cases involving *indicative conditionals*.

For cases of this kind, see, e.g., Goldstein and Santorio (2021), Ciardelli and Ommundsen (2022), Fusco (2022), or McNamara and Zhang (MS).

Some Bayesians lament this:

> [Bayesians] have no clear conception of what it might be to [update] on a conditional. (Skyrms, 1980, p. 169)

> [U]pdating on conditionals [seems to be] very different from standard Bayesian updating. (Douven, 2012, p. 240)

> Although [indicative conditionals] appear to play a central role in logical and uncertain reasoning... the relationship between [them] and the norms of Bayesian epistemology remains largely opaque. (Eva et al., 2019, p. 461).

My Goal: Show that the Bayesian update rules give the right answers in apparently problematic cases involving indicative conditionals.

Strategy: Appeal to a *sequence semantics* for conditionals, largely inspired by van Fraassen (1976).

Ground-clearing: I'll assume that conditionals have *truth-conditions* (contra expressivists). And I'll focus exclusively on *simple* conditionals in what follows—i.e., conditionals that have non-conditional propositions as antecedents consequents.

For ease of exposition, I'll focus mostly on a single, much-discussed case involving indicative conditionals—namely, van Fraassen's own *Judy Benjamin problem* (van Fraassen, 1981).

1 The *Judy Benjamin* Problem

> *Judy Benjamin*. Judy Benjamin, a soldier, has just been dropped into unfamiliar territory with her platoon. The territory is divided into two halves: Red Territory (R) and Blue Territory ($\neg R$). Each of these halves is divided into two further halves: Headquarters Company Territory (H), and Second Company Territory ($\neg H$). Initially, Judy is equally confident that she's in each of the four quadrants. But after some time, her captain appears on the radio, and says the following:

To say that Judy is equally confident in each of the four quadrants is to say that her credences are such that $p(R \wedge H) = 1/4$, $p(R \wedge \neg H) = 1/4$, etc.

(1) The probability is 3/4 that if you're in Red Territory, then you're in Headquarters Company Territory."

> The radio then crackles and dies. How should Judy's credences change when she hears the Captain utter the indicative conditional (1)?

Desiderata (van Fraassen, 1981):

I'm also going to assume that Judy's posterior credences should be such that $q(R \to H) = 3/4$.

(J1) Judy's new conditional credence in H given R should be 3/4: $q(H \mid R) = 3/4$.

(J2) Judy's unconditional credence in R should remain unchanged: $p(R) = q(R) = 1/2$.

(J3) For any proposition and any $X \in \{R \wedge H, R \wedge \neg H, \neg R\}$: $p(- \mid X) = q(- \mid X)$.

The idea underpinning (J3) is that the Captain's utterance or (1) shouldn't cause any change in Judy's conditional credences, except the change in $p(H \mid R)$ and thus also $p(\neg H \mid R)$, respectively.

Van Fraassen claims that if Judy's posterior credences satisfy (J1)–(J3), then she can't be updating in accordance with the standard Bayesian update rules (viz., conditionalization/Jeffrey conditionalization).

Actually, van Fraassen claims something more general—namely, that if Judy satisfies (J1)–(J3), then she can't be udpating in a way involves any plausible form of "distance minimization" between her prior credence function and her posterior credence function. I'll focus on the special case of (Jeffrey) conditionalization here.

2 Two Common Responses

There are two standard responses to *Judy Benjamin* in the literature:

(i) We should deny some of van Fraassen's desiderata.

(ii) We should introduce additional updating rules, which supplement standard Bayesianism.

Most authors who make the first type of response focus on (J2). In fact, van Fraassen himself denies that desideratum in later work.

See van Fraassen et al. (1986). See also, e.g., Joyce (2004), Boven (2009, Vasudevan (2020).

To see why, consider an altered version of (1):

(2) If you're in Red Territory, then you're in Headquarters Company Territory.

So, rather than telling Judy there's merely a high probability of being in H-territory, if she's in R-territory, the Captain now tells her that this is a certainty.

Adapting (J1) accordingly:

(J1*) Judy's conditional credence in H given R should be 1: $q(H \mid R) = 1$.

But $q(H \mid R) = 1$ implies that $q(R \wedge \neg H) = 0$. In turn, this implies that $q(R \supset H) = 1$ (provided Judy doesn't learn anything stronger than this). So, it looks like Judy should update her credences by conditionalizing on the *material conditional*, $R \supset H$.

To spell this out: suppose $q(H \mid R) = 1$. Then: $q(R \wedge \neg H) = 0$. This implies $q(\neg R \vee H) = 1$. But $\neg R \vee H$ is truth-functionally equivalent to $R \supset H$.

If that's right, then there are strong reasons to doubt (J2). After all, in general, $q(R) = p(R \mid R \supset H) \leqslant p(R)$ (this was proved by Popper and Miller (1983)). Extrapolating: there's some reason to think that Judy should *lower* her credence in R after hearing the Captain say (1).

Eva et al. (2019): "When we learn [an] indicative conditional, the antecedent becomes more informative (and hence more easily falsifiable) and less probable" (p. 468)

However, this view leaves a number of things to be explained:

(a) It's almost universally agreed that indicative conditionals are *not* material conditionals. So, why would their "learnability" conditions be the same?

However, on this point, see Santorio (2022).

(b) The "material conditionalization" view doesn't generalize easily to "uncertain" learning situations, where Jeffrey conditionalization usually applies.

(c) It just looks wrong to say, that *whenever* you learn a conditional, your credence in its antecedent should go down.

For example, see the "Ski Trip" example in Douven and Romeijn (2011). In that case, it looks strongly like your credence in the antecedent of the relevant conditional should go up!

So, what about responses of type (ii)—i.e., the response which says we should supplement Bayesianism with additional update rules?

Bradley (2005) proposes this rule:

Adams Conditionalization:

$$q(-) = p(A \wedge C \wedge -) \cdot \frac{q(C \mid A)}{p(C \mid A)} + p(A \wedge C \wedge \neg -) + \frac{q(\neg C \mid A)}{p(\neg C \mid A)} + p(\neg A \wedge -).$$

This rule looks pretty gnarly! But it turns out that, if Judy updates her credences in accordance with it, then all of (J1)–(J3) are satisfied.

Also, Adams conditionalization has a number of nice properties. As Bradley points out, for example, it's in some sense the precise converse of Jeffrey conditionalization. Whereas, in Jeffrey conditionalization, your unconditional credences change, and your conditional credences stay rigid, in Adams conditionalization it's the reverse.

Unfortunately, however, defenders of Adams conditionalization also leave a few questions unanswered:

(a) Even proponents of Adams conditionalization deny that it applies to *all* cases in which you learn an indicative conditional. For instance, Douven and Romeijn (2011) say:

> We are inclinded to think that Adams [conditionalization]... covers most of the cases of learning a conditional. Unfortunately, however, it would be wrong to think it covers all of them. (p. 654)

(b) Additionally, if indicative conditionals have truth-conditions, then why do we need an update rule that looks so dramatically different from the ordinary Bayesian rules?

This might be easy to explain on the expressivist view, where 'If A, then C' just expresses your conditional credences. But we're setting that view aside for present purposes.

3 Stalnaker's Thesis and Triviality

Intuitively, there's a connection between (J1)–(J3) and the thesis known as *Stalnaker's Thesis* (Stalnaker, 1970):

Stalnaker's Thesis. $p(A \to C) = p(C \mid A)$.

Versions of Stalnaker's thesis have been widely defended in the literature. To see why, consider this example:

(3) If I roll this fair, six-sided die, then it will land on 2.

How confident are you in that sentence? Presumably your answer is 1/6—which is just what Stalnaker's thesis requires.

For example, consider (J1). Eva et al. (2019) say that (J1) seems to be

> justified by the influential idea, commonly referred to as ['Stalnaker's thesis'], that the probability of the indicative conditional 'If A, then C' is given by the corresponding conditional probability $p(C \mid A)$. (p. 464, with trivial changes of notation)

Additionally, (J2) seems to require that Judy's credence in R be *independent* of her credence in $R \to H$:

$$p(R) = p(R \mid R \to H).$$

However, it's easy to show that this sort of "antecedent independence" *entails* Stalnaker's thesis, given some background assumptions about indicative conditionals.

The main assumption we need is what's called *probabilistic centering*: $p(A \wedge (A \to C)) = p(A \wedge C)$. This assumption, together with antecedent independence, entails Stalnaker's thesis:

$$\begin{aligned} p(A \to C) &= p(A \to C \mid A) \\ &= p(A \wedge (A \to C))/p(A) \\ &= p(A \wedge C)/p(A) \\ &= p(C \mid A). \end{aligned}$$

Probabilistic centering falls out of the "sequence semantics" that I adopt below.

The parallels between (J1)–(J3) and Stalnaker's thesis partly explain, I think, why the former are so hard to accommodate within a standard Bayesian framework. After all, Stalnaker's thesis is itself very difficult to accommodate within that framework. The issue is that the thesis is subject to various *triviality results*, given some assumptions.

Example: *The Wallflower Problem* (Hájek, 1989):

(4) If I don't roll a 1 with this (fair, six-sided) die, then I'll roll a 2.

The first triviality results for Stalnaker's thesis were proved, famously, by Lewis (1976). There are literally *dozens* of other triviality results in the literature now.

Stalnaker's thesis says that your credence in (4) should be 1/5 (assuming your credence in each world w_i, where the die lands on i, is 1/6).

But suppose propositions are sets of possible worlds. Then, there can be no proposition composed of $w_1, ..., w_6$ whose credence is equal to

1/5, since any such proposition must get credence equal to some multiple of 1/6.

Thus, it seems like Stalnaker's thesis must be wrong. In turn, this gives us reasons to doubt that van Fraassen's desiderata (J1)–(J3) are correct.

4 Sequence Semantics

Saying that Stalnaker's thesis is wrong has been the most common reaction to the triviality results in the literature. But it isn't the only reaction.

Recently, a number of philosophers have argued that, in cases like the Wallflower example above, a better reaction is to say that we've *modelled* the situation incorrectly: six, coarse-grained worlds aren't sufficient to capture all the relevant (epistemic) possibilities.

How, then, *should* we model the situation?

Van Fraassen (1976) introduces a fine-grained model of conditional contents, based on Stalnaker's (1968) semenatics. Here's (approximately) how it works.

I present things a *little* differently here to how van Fraassen does. In particular, my presentation draws a lot on Goldstein and Santorio (2021) as well.

Start with a (finite) set of "factual" worlds, $\mathcal{W} = \{w_1, ..., w_n\}$. Given $\mathcal{W}$ we then generate a set of *sequences* of worlds. For example, if $\mathcal{W} = \{w_1, w_2, w_3\}$, then the set of all sequences is:

$$\mathcal{S} = \begin{Bmatrix} \langle w_1, w_2, w_3\rangle, \langle w_1, w_3, w_2\rangle, \\ \langle w_2, w_1, w_3\rangle, \langle w_2, w_3, w_1\rangle, \\ \langle w_3, w_1, w_2\rangle, \langle w_3, w_2, w_1\rangle \end{Bmatrix}.$$

In the van Fraassen-inspired framework, sequences function as the *points of evaluation* for indicative conditionals. In particular:

(i) a "factual" proposition, A, is *true at a sequence*, s, just in case it's true at the *first* world in the sequence; and

(ii) a conditional $A \rightarrow C$ is true at a sequence, s, just in case the first A-world in the sequence is a C-world.

So, we're thinking of worlds here as entities that pin down all the "descriptive facts", but not necessarily all the conditional (or modal) facts. This is akin to Stalnaker's idea, that a conditional's truth-value depends, not only on the facts that obtain at a world, but also on the choice of a selection function.

Thus, in the example above, suppose $A = \{w_1, w_2\}$ and $C = \{w_2, w_3\}$. Then $A \rightarrow C$ is true at $\langle w_3, w_2, w_1\rangle$, but not at $\langle w_3, w_1, w_2\rangle$.

We can constrain the "admissible" sequences in a context by introducing a background partition (Kaufman, 2004; Khoo, 2016), or an accessibility relation (Mandelkern, forthcoming). I won't say much about that here, however.

5 Credences and Updating

In the standard Bayesian framework, we assume that your credence function, p, is defined over a set of worlds, $\mathcal{W}$.

However, since we're now working with *sequences* of worlds, we need a way of "extending" your credence function, from worlds to sequences.

Cf. van Fraassen (1976), Khoo and Santorio (2018), Goldstein and Santorio (2021), Khoo (2022), and Mandelkern (forthcoming).

Thus, suppose we start with a credence function, p, defined over $\mathcal{W}$. Then, we can extend p to your "full" credence function over sequences, using a recursive procedure:

In this definition, $[w_1, ..., w_k]$ is the set of sequences beginning with $w_1, ...w_k$ (in that order).

(i) $p([w]) = \mathsf{p}(w)$;

(ii) $p([w_1, ..., w_k]) = p([w_1, ..., w_{k-1}]) \cdot \mathsf{p}(w_k \mid \mathcal{W} - \{w_1, ..., w_{k-1}\})$.

Heuristically: your credence in the sequence $\langle w_1, ..., w_n \rangle$ is your credence that you'll draw those worlds from an urn, in that order and without replacement.

Note also two interesting things. First, the function p, as we've defined it, preserves the credences p assign to "factual" propositions (and thus it preserves p's conditional credences, too).

Second, in light of this, we have the following result:

Theorem 1 (van Fraassen, 1976; Goldstein and Santorio, 2021; Khoo, 2022; Mandelkern, forthcoming)**.** Suppose that p comes from p by the procedure (i)–(ii). Then, for all factual A, C, $p(A \rightarrow C) = p(C \mid A)$. That is, Stalnaker's thesis holds.

Thus, in the sequence-based framework, Stalnaker's thesis can be satisfied non-trivially—contra the triviality results of Lewis and others.

Next, we need to say how updating on conditionals works in this new setting. Here's what I propose—it's basically what you'd expect:

The definitions below require me to endorse Conditional Excluded Middle (CEM). But that falls out naturally of the sequence-based semantics for indicative conditionals sketched above.

(i) **Conditionalization**. After learning $A \rightarrow C$ with certainty, $q(-) = p(- \mid A \rightarrow C)$.

(ii) **Jeffrey Conditionalization**. Suppose $q(A \rightarrow C) \neq p(A \rightarrow C)$. Then:

$$q(-) = p(- \mid A \rightarrow C) \cdot p(A \rightarrow C) + p(- \mid A \rightarrow \neg C) \cdot p(A \rightarrow \neg C).$$

So, in effect this says that, after you have a learning experience involving an indicative conditional, you should update in the ordinary Bayesian way—but this time, the content of what's learned is a set of sequences (rather than just a set of worlds).

6 Back to *Judy Benjamin*

It turns out that—given only one assumption—the semantic theory sketched above, together with the view of updating just described, satisfies all of van Fraassen's desiderata in *Judy Benjamin*.

I'll illustrate this here with a simplified version of that case.

The "full" version of the case is worked out in my paper. I'm happy to share it, if you're curious.

Let Judy's epistemically possible worlds be $\mathcal{W} = \{w_1, w_2, w_3, w_4\}$. In particular, the epistemic possibilities Judy countenances are summarized in the following table:

Table 1: Judy's Epistemic Possibilities

	H	$\neg H$
R	w_1	w_2
$\neg R$	w_3	w_4

Now, suppose Judy hears the Captain utter (1), and updates from p to q by Jeffrey conditionalization on the set of all sequences. Then:

(i) (J1) is satisfied straightforwardly. After all, if $q(R \to H) = 3/4$, then it follows from Theorem 1 that $q(H \mid R) = 3/4$.

(ii) Now consider (J2). Before, we said that antecedent independence entails Stalnaker's thesis, given probabilistic centering. But it turns out that Stalnaker's thesis and probabilistic centering also entail antecedent independence. So, in virtue of the fact that $q(R \to H) = q(H \mid R)$, it follows that $q(R \mid R \to H) = q(R)$. In other words, changes in Judy's credence in $R \to H$ don't affect her credence in R. So, $q(R) = p(R) = 1/2$.

Here's a proof that Stalnaker's thesis and probabilistic centering entail antecedent independence:

$$\begin{aligned} p(A \to C) &= p(C \mid A) \\ &= p(A \wedge C)/p(A) \\ &= p(A \wedge (A \to C))/p(A) \\ &= p(A \to C \mid A) \end{aligned}$$

(iii) Now turn to (J3). Recall that this says that Judy's updated credence in any proposition, conditional on $X \in \{R \wedge H, R \wedge \neg H, \neg R\}$ should be such that $q(- \mid X) = p(- \mid X)$.

- In the first case, $R \wedge H = \{w_1\}$. And of course, $q(- \mid w_1) = p(- \mid w_1) \in \{0, 1\}$. So that case is easy.
- The second case is the same, since $R \wedge \neg H = \{w_2\}$.
- Finally, checking the case of $\neg R$ involves some tedious calculations. But in that case, too, we get that $q(- \mid \neg R) = p(- \mid \neg R)$.

Again, these calculations are in my paper, which I'm happy to share.

Thus, all of desiderata (J1)–(J3) are satisfied. And this is so, even though we've assumed that Judy updates by Jeffrey conditionalization.

7 A More General Result

Updating by Jeffrey conditionalization on a set of sequences gives the same results as Adams conditionalization in the *Judy Benjamin* problem. This might lead you to wonder whether there's any systematic connection between the two.

There is. In my paper, I prove the following result:

An even more general version of this result was proved independently by Snow Zhang. Instead of relying on a particular semantics, Snow shows that the result holds for any semantics that satisfies some abstract conditions.

Theorem 2. Let $\mathcal{W}$ be a set of possible worlds, and let $\mathcal{S}$ be the set of all sequences we can generate from $\mathcal{W}$. Let p be a probability function defined on $\mathcal{W}$, and let p be a credence function "lifted" from p according to (i) and (ii). Then, updating by Jeffrey conditionalization on $\{A \to C, A \to \neg C\}$ is equivalent to Adams conditionalization on $\{A \wedge C, A \wedge \neg C, \neg A\}$.

For example, in transparent contexts, E guarantees that Independence holds.

So, in a strong sense, Adams conditionalization just *is* Jeffrey conditionalization (albeit in a more fine-grained space).

[References available upon request.]