Updating for Externalists

J. Dmitri Gallow

The character I'll call 'the internalist' holds that your evidence can never fail to tell you what your total evidence is. If your total evidence is e, then you must have the evidence that your total evidence is e. The character I'll call 'the externalist' denies this (§1). An *update* is a strategy for revising your degrees of belief, or *credences*, in response to the outcome of an experiment. The internalist has their update: upon learning e, adopt your pre-experimental credences conditional on e. This is the rule of conditionalization. Salow (forthcoming) teaches that if the externalist adopts conditionalization, then they will be capable of engaging in deliberate self-delusion—designing experiments which are guaranteed to raise their credence that ϕ as high as they like, independent of whether ϕ is true or false (§3). This is not rational inquiry, and no sensible epistemology will call it such. The externalist should reject conditionalization. So the externalist is in need of an update. I have one to offer (§4). This update has maximal expected accuracy amongst those which the externalist should regard as genuinely available. In experiments where the internalist has it right and your evidence will tell you what your total evidence is, this update agrees with conditionalization. It similarly agrees with the updates of Jeffrey (1965), Hild (1998a,b), Schoenfield (forthcoming), and Gallow (2014) in the paradigm cases for which those updates were designed.

1 Internalism and Externalism

In general, to be an internalist is to think that some condition is within our epistemic reach. To be an externalist is to deny this. Given some condition c, an internalist claims that, if you satisfy c, then you have access to the fact that you satisfy c. And an externalist believes that you can satisfy condition c without having access to the fact that you satisfy c. Different conditions and different kinds of access yield different breeds of internalism and externalism. Let the condition be possessing the total evidence e. Say that you have access to a fact iff it is entailed by your evidence. Then, the internalist says: whenever e is your total evidence,

Draft of October 15, 2018; Word count: 11,443
Comments appreciated. ✉: jdmitrigallow@pitt.edu

your evidence entails that e is your total evidence. The externalist: on the contrary, sometimes e can be your total evidence without you having the evidence that e is your total evidence.

> Internalism
> Necessarily, if e is your total evidence at t, then at t you have the evidence that e is your total evidence at t
>
> $$\Box(\mathbb{T}_t e \rightarrow \mathbb{E}_t \mathbb{T}_t e)$$

> Externalism
> Possibly, you have the total time t evidence e without having the evidence at t that e is your total time t evidence.
>
> $$\Diamond(\mathbb{T}_t e \wedge \neg \mathbb{E}_t \mathbb{T}_t e)$$

'$\mathbb{T}_t e$' denotes the proposition that e is your total evidence at t. '$\mathbb{E}_t e$', that e is entailed by your evidence at t.

The plausibility of internalism and externalism will depend in part upon our conception of evidence. My primary focus here will be on the way that these theses interact with the principle of conditionalization (to be introduced below). Conditionalization says to become certain that your evidence is true. For this reason, throughout, I will understand '$\mathbb{E}_t e$' as saying at least that, at t, it has become rational for you to be certain that e. And I will understand '$\mathbb{T}_t e$' as saying at least that, at t, it has become rational for you to be certain that e, and it has not become rational for you to be certain about anything stronger than e. Beware: others use 'evidence' differently.[1] Insofar as others utilize a conception of evidence on which your total evidence can be e without it being rational for you to be certain that e, or one on which you can be certain of ϕ even when ϕ is not entailed by your total evidence, the theses they dub 'internalism' and 'externalism' may differ from the ones I am discussing. To have a name, call a proposition about which certainty has been rationalized 'certainty evidence'. In §§1–3, whenever I say 'evidence', I refer to certainty evidence. In §4, I consider other notions of evidence and their relationship to certainty evidence.

1.1 Internalism and Externalism with Kripke Frames

We may provide a standard Kripke semantics for the operators $\mathbb{E}_t$ and $\mathbb{T}_t$. On that semantics, $\mathbb{E}_t$ is a familiar necessity modal—$\ulcorner \mathbb{E}_t e \urcorner$ is true at w iff $\ulcorner e \urcorner$ is true at all worlds to which w bears a time t accessibility relation, R_t. $\mathbb{T}_t$ is less

[1] In particular, Hild (1998b,a), Williamson (2000), Gallow (2014), and Schoenfield (forthcoming). See §4.

familiar, but its semantics are simple enough: $\ulcorner \mathbb{T}_t e \urcorner$ is true at w iff $\ulcorner e \urcorner$ is true at *all and only* worlds to which w bears R_t. (Going forward, I suppress time indices.) If we assume that evidence is factive, then the internalist's thesis, $\mathbb{T}e \to \mathbb{E}\mathbb{T}e$, is equivalent to the S5, or 'negative access', principle $\neg\mathbb{E}e \to \mathbb{E}\neg\mathbb{E}e$. S5 says that the lack of evidence is always evidence itself. It is equivalent to the conjunction of the S4, or 'positive access', principle $\mathbb{E}e \to \mathbb{E}\mathbb{E}e$ and the Brouwer ('B') principle $\neg\mathbb{E}\neg\mathbb{E}e \to e$. S4 says that the possession of evidence is always evidence itself. B says that any evidence you might have, for all your evidence has to say, is true.

Williamson teaches that cases of perceptual illusion counterexample B. In the bad case, you look at a white wall illuminated with red lighting. In the good case, you look at a red wall illuminated with white lighting. In the bad case, it is false that the wall is red, but your evidence does not rule out that you are in the good case. In the good case, you have the evidence that the wall is red. So, in the bad case, your evidence does not rule out that you have the evidence that the wall is red. Even so, the wall is not red. So $\neg\mathbb{E}\neg\mathbb{E}r \wedge \neg r$, where '$r$' is that the wall is red. So B is false. The internalist will deny that in the good case you have the evidence that the wall is red. Rather, in both the good and the bad case, you merely have the evidence that the wall appears red, and/or that you believe the wall is red. B is salvaged—though skepticism looms.

Williamson similarly teaches that cases in which our perceptual knowledge is inexact counterexample the 'postive access' principle S4. Off in the distance, you catch a glimpse of an unmarked clock (see figure 1a). Your vision is good enough for you to get the evidence that the hand is on the right-hand side of the clock. And though you likely learn something stronger still, you don't learn the precise location of the clock hand.[2] At most, you learn that the clock hand is located in some interval (see figure 1b). Grant also that, if the clock hand is located at a position b, then you won't learn that the clock hand is located within some interval that has b as an endpoint (see figure 1c). Grant not only that this is true, but that you've learned it.

These assumptions contradict S4.[3] For the following three claims are inconsistent (In the following, I use 'H' as a variable for the position of the clock hand).

A1) The strongest proposition you learn about the position of the clock hand is that it lies in some interval $[a, b]$, with $a < b$.

A2) You have the evidence that, if the clock hand is located at b, then your

[2] Throughout, I will use 'learn that e' to mean 'acquire the evidence e'.

[3] We must also take on board the following assumptions, both of which are presupposed by the usual possible worlds semantics: 1) the evidence operator $\mathbb{E}$ satisfies the K-axiom; *i.e.*, $\mathbb{E}(\phi \to \psi) \to \mathbb{E}\phi \to \mathbb{E}\psi$; and 2) if ϕ is evidence and ψ is logically equivalent to ϕ, then ψ is evidence as well.

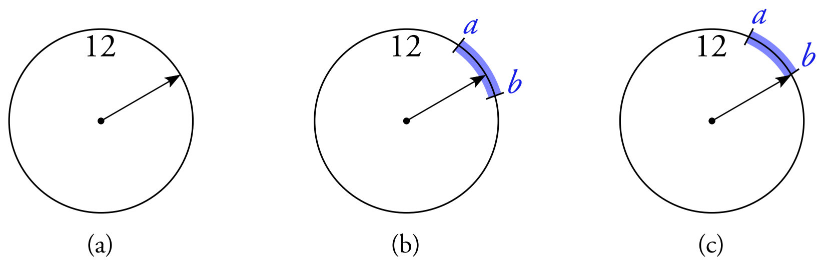

FIGURE 1: A distant and brief glimpse at the unmarked clock (1a) provides the evidence that the clock hand is positioned within some interval of values $[a, b]$ (1b); and you've learned that, if the clock hand is positioned at b, then the glimpse does not provide the evidence that the clock hand is no further than b (1c). These assumptions contradict the 'positive access' principle for evidence, $\mathbb{E}\phi \to \mathbb{E}\mathbb{E}\phi$.

evidence doesn't tell you that it is located no further than b.

$$\mathbb{E}[H = b \ \rightarrow \ \neg\mathbb{E}(H \leqslant b)]$$

A3) The possession of evidence is always evidence itself.

$$\mathbb{E}\phi \to \mathbb{E}\mathbb{E}\phi$$

By contraposition on (A2),

(B1) $$\mathbb{E}\,[\mathbb{E}(H \leqslant b) \ \rightarrow \ H \neq b]$$

Assuming that the evidence operator $\mathbb{E}$ satisfies the K-axiom, (B1) entails (B2).

(B2) $$\mathbb{E}\mathbb{E}(H \leqslant b) \to \mathbb{E}\,(H \neq b)$$

(A1) entails (B3).

(B3) $$\mathbb{E}(H \leqslant b)$$

From (B3) and (A3), we have

(B4) $$\mathbb{E}\mathbb{E}(H \leqslant b)$$

From (B4) and (B2),

(B5) $$\mathbb{E}(H \neq b)$$

(B5) contradicts (A1), which assured us that the *strongest* thing you learned about the position of the clock hand was that it was within the interval $[a, b]$. Since this does not entail $H \neq b$, (A1) tells us that you cannot have learned it.[4]

So (A1), (A2), and (A3) are inconsistent. Williamson thinks that (A3), the positive access principle, is the least plausible of the three; but others, like Salow (forthcoming) and Stalnaker (2009), choose instead to reject (A2) and retain the positive access principle.[5]

1.2 Internalism and Externalism with Experiments

(A1) and (A2) are enough to counterexample S4, but we could make do with less. Say that the strongest thing you stand to learn is some proposition about the position of the clock hand. Go on to say that you may learn that it lies in the interval I_1 (and no more); and likewise, you may learn that it lies in the interval I_2 (and no more). Add that I_1 and I_2 overlap, and you will have contradicted S4. For if you are in a position to learn one of two overlapping evidence propositions (and no more), then internalism's S5 principle is false. If one potential piece of evidence entails the other, then B is false. Else, S4 is false.

Call a set of propositions $\mathcal{E} = \{e_1, e_2, \ldots, e_N\}$ an *experiment*. Say you are *conducting the experiment* $\mathcal{E}$ at time t iff the following is true of you.

C1) Your time t total evidence might be e_1.

C2) Your time t total evidence might be e_2.

$\vdots$

CN) Your time t total evidence might be e_N.

C$N+1$) It must be that: either your time t total evidence is e_1, or your time t total evidence is e_2, or $\ldots$, or your time t total evidence is e_N.

This is a broad notion of 'conducting an experiment'. All it takes to conduct an experiment in this sense is for there to be a set of propositions you might come to learn at t. Opening a drawer to find pens, checking the front page of the New York Times, and looking at your wristwatch could all count as conducting an experiment in this sense.[6]

---

[4] The reader may be wondering whether this contradiction may be avoided by exchanging (A1)'s closed interval $[a, b]$ for an open one (a, b)—call the resulting claim '(A1*)'. (A1*) will be inconsistent with and (A3) and the following principle, for any choice of $\epsilon > 0$, no matter how small—the reasoning is exactly the same as in the body, *mutatis mutandis*:

$$\mathbb{E}[H = b - \epsilon \rightarrow \neg\mathbb{E}(H < b)]$$

[5] Stalnaker (2009) does not explicitly discuss the positive access principle for *evidence*; his focus is S4 for *rational belief*.

[6] I borrow terminology from Greaves & Wallace (2006).

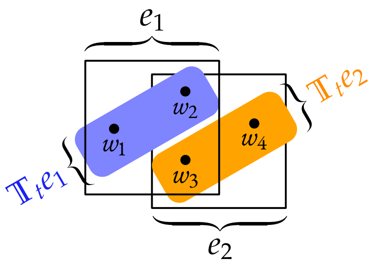

FIGURE 2: There are four epistemically possible worlds, $\{w_1, w_2, w_3, w_4\}$. You conduct the experiment $\{e_1, e_2\}$ at t, where $e_1 = \{w_1, w_2, w_3\}$, $e_2 = \{w_2, w_3, w_4\}$, $\mathbb{T}_t e_1 = \{w_1, w_2\}$, and $\mathbb{T}_t e_2 = \{w_3, w_4\}$. Though $\{\mathbb{T}_t e_1, \mathbb{T}_t e_2\}$ is a partition, $\{e_1, e_2\}$ is not.

By definition, if you conduct the experiment $\mathcal{E} = \{e_1, e_2, \ldots, e_N\}$ at time t, $\{\mathbb{T}_t e_1, \mathbb{T}_t e_2, \ldots, \mathbb{T}_t e_N\}$ is a partition.[7] The definition leaves open whether $\mathcal{E}$ *itself* is a partition. Consider figure 2. You will either learn e_1 (and no more) or e_2 (and no more), and e_1 and e_2 are consistent. Even though $\{\mathbb{T}_t e_1, \mathbb{T}_t e_2\}$ forms a partition, $\{e_1, e_2\}$ does not. The internalist insists that, necessarily, any experiment you conduct is a partition. This is equivalent to their thesis. The following are all equivalent,

D1) For all e and all t, necessarily, $\mathbb{T}_t e \to \mathbb{E}_t \mathbb{T}_t e$.

D2) For all e and all t, necessarily, $\neg\mathbb{E}_t e \to \mathbb{E}_t \neg\mathbb{E}_t e$.

D3) For all e and all t, necessarily, both $\mathbb{E}_t e \to \mathbb{E}_t \mathbb{E}_t e$ and $\neg\mathbb{E}_t \neg\mathbb{E}_t e \to e$.

D4) For all t, necessarily, if you conduct the experiment $\mathcal{E}$ at t, then $\mathcal{E}$ is a partition.

The internalist accepts (D1–D4), while the externalist accepts their negations, (E1–E4).

E1) For some e and some t, possibly $\mathbb{T}_t e \wedge \neg\mathbb{E}_t \mathbb{T}_t e$.

E2) For some e and some t, possibly $\neg\mathbb{E}_t e \wedge \neg\mathbb{E}_t \neg\mathbb{E}_t e$.

E3) For some e and some t, possibly either $\mathbb{E}_t e \wedge \neg\mathbb{E}_t \mathbb{E}_t e$ or $\neg\mathbb{E}_t \neg\mathbb{E}_t e \wedge \neg e$.

E4) For some t, possibly you conduct a non-partitional experiment $\mathcal{E}$ at t.

There are two ways of understanding the internalist thesis, corresponding to two different readings of the 'necessarily's in (D1–D4). We could read them as either metaphysical or epistemic necessity modals. On the metaphysical reading, internalism says that it is metaphysically impossible for you to ever possess total

[7] For our purposes, a *partition* is a set of mutually exclusive and jointly exhaustive propositions—a set of propositions exactly one of which must be true (read the 'must' as epistemic).

evidence e without possessing the evidence that e is your total evidence. On the epistemic reading, internalism says that you will always be able to rule out, in advance, that you will acquire total evidence e without also acquiring the evidence that e is your total evidence. The corresponding readings of the 'possibly's in (E1–E4) give us epistemic and metaphysical flavors of externalism.

Both forms of internalism and externalism are interesting, but I will confine attention to the epistemic versions of the views. My topic is updating strategies, and your update strategy should depend upon what is epistemically possible for you. Suppose—perhaps *per impossibile*—that, while it is metaphysically possible for $\mathbb{T}e$ to be true without you learning it, this is not epistemically possible. Then you should not plan for the contingency in which you learn e without also learning $\mathbb{T}e$. From your benighted perspective, this contingency is impossible.

Our disputants are therefore debating about the propriety of certain prospective epistemic states. Is it rationally permissible for you to foresee possibilities in which you get the total evidence e without getting the evidence $\mathbb{T}e$? The internalist says: no, definitely not. The externalist: yes, perhaps.

1.3 Introspective Evidence

The internalist places restrictions of the kinds of evidence you can foresee yourself receiving. These restrictions can appear unduly strong. Consider the following case.

> SNEAK PEEK
>
> You and Bonnie are playing a game involving three cups and a ball. While your back is turned, Bonnie places the ball under one of the cups and shuffles them around. When she's done, you will attempt to guess which cup hides the ball. If you guess correctly, you win; if not, Bonnie wins. You're certain that there's no funny business, so that the ball is either beneath cup 1, cup 2, or cup 3 (call these propositions '1', '2', and '3', respectively). At t, while the cups are being shuffled, your accomplice is going to attempt to distract Bonnie, and you're going to sneak a peek under the cup closest to you at t, if you can. However, you don't know which cup will be closest to you at t, and therefore, you don't know which cup you'll try to look under. Nor do you know whether you'll be successful.

In SNEAK PEEK, it is reasonable to think that 1, 2, and 3 are each equally likely. *Prima facie*, you might acquire any of the following total evidence propositions:

F1) Nothing at all ($\top$), which you will learn if you aren't able to sneak a peek at t.

F2) The ball is not beneath cup 1 ($\neg 1$), which you will learn if you look beneath cup 1 at t and don't see the ball.

F3) The ball is not beneath cup 2 ($\neg 2$), which you will learn if you look beneath cup 2 at t and don't see the ball.

F4) The ball is not beneath cup 3 ($\neg 3$), which you will learn if you look beneath cup 3 at t and don't see the ball.

F5) The ball is beneath cup 1 (1), which you will learn if you look beneath cup 1 at t and see a ball.

F6) The ball is beneath cup 2 (2), which you will learn if you look beneath cup 2 at t and see a ball.

F7) The ball is beneath cup 3 (3), which you will learn if you look beneath cup 3 at t and see a ball.

If that's right, then the experiment you are conducting at t does not form a partition. Since it is clearly possible to conduct this experiment, doesn't SNEAK PEAK suffice to establish externalism?

It does not. The internalist should grant that each of (F1–F7) could be the strongest thing you learn *about the location of the ball.* However, if you are rational, you will be certain to receive more evidence than this. For instance, you will be certain to also receive the evidence of how your credences have changed in response to this evidence about the location of the ball. Call evidence of this kind 'introspective evidence'. Even though the consistent propositions $\neg 1$ and $\neg 2$ both could be the strongest propositions you learn about the location of the ball, neither $\neg 1$ nor $\neg 2$ could be the strongest proposition you learn *full stop*. If you learn that the ball isn't under cup 1, and you are rational, then you will become certain that $\neg 1$, and you will, moreover, learn that you have become certain that $\neg 1$. If you learn that the ball isn't under cup 2, and you are rational, then you must also learn that you've become certain that the ball isn't under cup 2. And you are certain in advance that you won't both become certain that the ball isn't under cup 1 and that the ball isn't under cup 2. Once your introspective evidence is taken into account, your experiment will form a partition after all.[8]

Our externalist wished to model your (pre-experimental) credal state in SNEAK PEEK with just three possibilities, 1, 2, and 3, each of which you took to be equally likely:

1	2	3
1/3	1/3	1/3

This representation of your pre-experimental credal state gives our externalist all they need to know in order to say how your credences about the location of the

[8] Introspective evidence is not the only kind of evidence an internalist could appeal to in order to justify their claim that, for all e, necessarily, $\mathbb{T}e \to \mathbb{E}\mathbb{T}e$. Still, this breed of internalism will be my focus here.

ball should change after the experiment. At least, it does so if we suppose the principle of CONDITIONALIZATION. As I will understand it here, CONDITIONALIZATION supposes that, prior to an experiment $\mathcal{E} = \{e_1, e_2, \ldots, e_N\}$, you should have a *strategy* for updating your credences in respponse to each potential evidence proposition $e \in \mathcal{E}$. For each $e \in \mathcal{E}$, use 'C_e' for the credence function you plan to adopt upon learning e and no more. CONDITIONALIZATION says: C_e should be your pre-experimental credence function, C, conditioned on the total evidence e, $C(- \mid e)$.[9]

> CONDITIONALIZATION
> If you conduct the experiment $\mathcal{E}$ with the pre-experimental credence function C, then, for each $e \in \mathcal{E}$, your strategy for responding to total evidence e, C_e, should be to condition C on e. That is, for all propositions ϕ,[10]
>
> (CONDI) $$C_e(\phi) \stackrel{!}{=} C(\phi \mid e)$$

By CONDI, if you were to get the evidence that the ball is not under cup 1, you should transition to this new credal state:

1	2	3
0	1/2	1/2

And, if you were to get the evidence that the ball is not under cup 2, you should transition to this new credal state:

1	2	3
1/2	0	1/2

While this representation serves the needs of our externalist perfectly well, it is not complete. For, if you are in the experiment $\mathcal{E} = \{1, 2, 3, \neg 1, \neg 2, \neg 3, \top\}$, then exactly one of the propositions in $\mathcal{E}$ will be your total evidence. If you have planned for this experiment, then, for each $e \in \mathcal{E}$, there is some credence function, C_e, you plan to adopt post-experiment if your total evidence is e. Let $\mathbf{U}e$ (for *u*pdate) be the proposition that you adopt the credence function C_e. A complete representation of your credal state should include these propositions as well. I assume you are certain to update to at least one of the credence functions

9 I will be taking for granted throughout that a rational credence function will be a probability. I'll also be making the simplifying assumptions throughout that a) the number of propositions over which C is defined is finite; and b) C assigns positive credence to every proposition compatible with your evidence.

10 I place an exclamation mark over the equals sign to indicate that the equality holds with normative, and not descriptive, force. CONDI claims not that your strategy for responding to e *will* be to condition on e, but rather that it *should* be.

	1	2	3
U1	6/42	0	0
U2	0	6/42	0
U3	0	0	6/42
U¬1	0	3/42	3/42
U¬2	3/42	0	3/42
U¬3	3/42	3/42	0
U⊤	2/42	2/42	2/42
	1/3	1/3	1/3

(a)

	1	2	3
U1	0	0	0
U2	0	6/28	0
U3	0	0	6/28
U¬1	0	3/28	3/28
U¬2	0	0	3/28
U¬3	0	3/28	0
U⊤	0	2/28	2/28
	0	1/2	1/2

(b)

	1	2	3
U1	0	0	0
U2	0	0	0
U3	0	0	0
U¬1	0	1/2	1/2
U¬2	0	0	0
U¬3	0	0	0
U⊤	0	0	0
	0	1/2	1/2

(c)

FIGURE 3: Figure 3b shows the result of conditioning the pre-experimental credence distribution from figure 3a on ¬1. Figure 3c shows the result of conditioning the pre-experimental credence distribution from figure 3a on **U**¬1.

C_e. For each $e \neq e^*$, C_e will be distinct from C_{e^*}, so you will not adopt more than one of these credence functions.[11] So $\{\mathbf{U}e_1, \mathbf{U}e_2, \ldots, \mathbf{U}e_N\}$ will form a partition. So there should be some possible world for each non-zero cell in the 3×7 grid in figure 3a. (I assume away possibilities at which $\mathbf{U}e$ but not e—rational certainty is factive, and you have chosen your plans accordingly. In the figure, I suppose that you are equally likely to update to each C_e.)

Introspective propositions like these are crucial for our internalist. According to them, if you learn that the ball is under cup 1, then you must additionally learn that you have updated to the credence function C_1, **U**1. By CONDI, if you stand to learn **U**1, then you must plan to become certain of it. The same goes for every other proposition about the position of the ball. So, while our externalist may think that your experiment is the non-partitional $\{1, 2, 3, \neg 1, \neg 2, \neg 3, \top\}$, our internalist will insist that it is instead the partitional {**U**1, **U**2, **U**3, **U**¬1, **U**¬2,

[11] For all e, e is your total evidence iff e is the strongest proposition about which certainty has become rational. So C_e should be certain that e and no more, and C_{e^*} should be certain that e^* and no more. Since $e \neq e^*$, C_e should not be equal to C_{e^*}.

U$\neg 3$, **U**$\top$}.

Suppose you are in the very center cell of the grid in figure 3a—fourth row, second column. The ball is under cup 2, you check cup 1, and find it empty. If your total evidence is just that the ball is not under cup 1—if you do not additionally learn that you've updated to $C_{\neg 1}$—then conditioning on your total evidence will take you to the post-experimental credence distribution shown in figure 3b. On the other hand, if you additionally acquire the introspective evidence that you've updated to $C_{\neg 1}$, then conditioning on this total evidence will take you to the post-experimental credence distribution shown in figure 3c.

Sneak peek is an overly simplistic example. Most externalists will think that you acquire some introspective evidence in cases like sneak peek. The cases in which they think we lack introspective access and conduct non-partitional experiments will be more complicated and more psychologically plausible. My goal is not to argue for externalism, but rather to develop it, so I'll continue to focus on simple examples like sneak peek. The lessons we learn there will carry over to more realistic cases.

2 Externalism, Conditionalization, and Reflection

In this section, I will argue that the externalist cannot accept both condi and the recommendations of van Fraassen (1984, 1995)'s principle of reflection in a particular kind of experiment. Moreover, the externalist cannot plausibly reject the recommendations of reflection in this experiment. I will conclude, then, that the externalist should reject condi.

Suppose that externalism is correct, and you are conducting the experiment $\mathcal{E} = \{\neg 1, \neg 3\}$, where $\{1, 2, 3\}$ is a partition and each cell is equally likely. Perhaps you're trying to find the ball in Bonnie's cups, you've guessed that it's under cup 2, and you know that Bonnie will reveal some empty cup you haven't guessed, either cup 1 or cup 3.[12] Then, you may find yourself in the pre-experimental credal state shown in figure 4. (If you are confused by rows 2 and 3 of this figure, they will be addressed below, and further discussed in §4.)

Suppose you are a conditionalizer, so that you are certain that you will either condition on $\neg 1$ or $\neg 3$. Notice: if you condition on $\neg 1$, your credence that 2 will rise to 1/2, since $C(2 \mid \neg 1) = 1/2$. And if you condition on $\neg 3$, your credence that 2 will rise to 1/2, since $C(2 \mid \neg 3) = 1/2$. So, prior to conducting the experiment, you are certain that your credence that 2 will rise to 1/2. Why wait? Cut to the chase—go ahead and adopt a credence of 1/2 in 2 before looking. Given what you know, you are in a position to rationally reason your way to adopting a credence of 1/2 in the proposition 2 before looking. If you are a conditionalizer, then you

[12] *Cf.* the 'Monty Hall' problem from Selvin (1975), and the 'three prisoners paradox' in Gardner (1961).

	1	2	3	
$\mathbf{U}\neg 1 \cap \mathbb{T}\neg 1$	0	1/12	4/12	5/12
$\mathbf{U}\neg 3 \cap \mathbb{T}\neg 1$	0	1/12	0	1/12
$\mathbf{U}\neg 1 \cap \mathbb{T}\neg 3$	0	1/12	0	1/12
$\mathbf{U}\neg 3 \cap \mathbb{T}\neg 3$	4/12	1/12	0	5/12
	1/3	1/3	1/3	

FIGURE 4: Your credal state before conducting the experiment $\mathcal{E} = \{\neg 1, \neg 3\}$.

should cut to the chase. But cutting to the chase is inconsistent with being a conditionalizer. For, if your credence remains at 1/2 after learning $\neg 1$, you will not have updated by conditioning on $\neg 1$. No matter your credence in 2, $C(2 \mid \neg 1) > C(2)$—and, likewise, no matter your credence in 2, $C(2 \mid \neg 3) > C(2)$.[13] If you are a conditionalizer, then you shouldn't be. So you shouldn't be.

Cutting to the chase is recommended by VAN FRAASSEN's principle of REFLECTION. As I'll understand it here, REFLECTION says that you should *defer* to your post-experimental self.[14] You defer to your post-experimental self just in case, for every proposition ϕ, your current credence that ϕ is your best estimate of the credence in ϕ which you will have post-experiment. Since C is a probability, your best estimates are given by your expectations. Suppose that your experiment is $\mathcal{E} = \{e_1, e_2, \ldots, e_N\}$, and you are certain that you will adopt one of the post-experimental credence functions in $\{C_{e_1}, C_{e_2}, \ldots, C_{e_N}\}$. As before, let $\mathbf{U}e$ be the proposition that, post-experiment, you update to the credence function C_e. Then, REFLECTION says:

> REFLECTION
> Your pre-experimental credence that ϕ should be equal to your expectation of the credence that ϕ you will have post-experiment,

(REFLECTION) $$C(\phi) \stackrel{!}{=} \sum_{e \in \mathcal{E}} C_e(\phi) \cdot C(\mathbf{U}e)$$

If you are a conditionalizer, then you violate REFLECTION in the experiment $\mathcal{E} = \{\neg 1, \neg 3\}$. You plan to raise your credence that 2 to 1/2 no matter what, so you expect your post-experimental credence that 2 to be 1/2, yet your pre-experimental

[13] Recall from footnote 9: I assume that $C(\phi) = 0$ iff ϕ is inconsistent with your evidence.

[14] VAN FRAASSEN's original principle enjoins you to defer to your future self, for all future times. Restricted to experiments which you are about to perform, in which you will not lose evidence, and for which you have an updating strategy, the principle I call 'REFLECTION' escapes many of the counterexamples to VAN FRAASSEN's principle (see BRIGGS (2009) for a nice taxonomy of those counterexamples). Those which remain involve credences *de se*. Credence *de se* will force us to reject or qualify REFLECTION and CONDITIONALIZATION both. Still, I'll ignore credence *de se* for the nonce.

credence that 2 is 1/3.

Perhaps your post-experimental self is not worthy of deference. In general, in externalist experiments, you will foresee possibilities in which your post-experimental credence is not the rational one to adopt, given your evidence—for instance, in the second and third rows of figure 4. (I will return to this feature of externalism in §4—see also Elga, 2013). And an *irrational* future version of yourself is not one deserving of deference. You should instead defer to your *rational* future self—that is, for each ϕ, your pre-experimental credence that ϕ should be your best estimate of the credence that ϕ which it would be *rational* for your future self to adopt, given their evidence. Call this principle RATIONAL REFLECTION.[15] The credence function C_e is the rational one to adopt iff your total evidence is e, $\mathbb{T}e$. So RATIONAL REFLECTION says:

RATIONAL REFLECTION

Your pre-experimental credence that ϕ should be equal to your expectation of your rational post-experimental credence that ϕ,

(RAT-REF) $$C(\phi) \stackrel{!}{=} \sum_{e \in \mathcal{E}} C_e(\phi) \cdot C(\mathbb{T}e)$$

If you are certain to update to C_e iff e is your total evidence, $C(\mathbf{U}e \leftrightarrow \mathbb{T}e) = 1$, say that you are *immodest*; else, say you are *modest*.[16] For instance, in figure 4, you are modest. In row 2, you foresee the possibility of updating on $\neg 3$ even when your total evidence is $\neg 1$. If you are immodest, then REFLECTION and RAT-REF will agree. In general, if you are modest, then REFLECTION and RAT-REF may come apart.

Even so: if CONDI is correct, then, when it comes to your credence that 2 in this experiment, both REFLECTION and RAT-REF speak with a single voice. Pre-experiment, you know that you will either receive the total evidence $\neg 1$ or $\neg 3$. If your total evidence is $\neg 1$, then CONDI counsels to raise your credence in 2 to 1/2. And CONDI says exactly the same thing if your total evidence is $\neg 3$. So, prior to conducting the experiment, you are certain that there is a fully rational agent—*viz.*, the rational version of your future self—who accepts your precise epistemic standards, who has strictly more evidence than you do, and whose credence that 2 is 1/2. Why wait? Cut to the chase—given what you know, you are in a position to rationally reason your way to adopting a credence of 1/2 in the proposition 2 before looking. If you should be a conditionalizer, then you should cut to the

---

15 See Christensen (2010). The principle discussed by Christensen is a purely *synchronic* principle requiring deference to your *currently* rational credences. Here, I generalize the principle to cover your future rational credences.

16 Beware: this terminology is idiosyncratic. Others will call you 'immodest' iff you are less than certain that your current degrees of belief are rational.

chase, but cutting to the chase is inconsistent with being a conditionalizer. No matter your credence that 2, $C(2 \mid \neg 1) > C(2)$ and $C(2 \mid \neg 3) > C(2)$. If you should be a conditionalizer, then you shouldn't be. So you shouldn't be.

What does this show? We might think that it shows the following claims to be inconsistent:[17]

G1) EXTERNALISM

G2) CONDITIONALIZATION

G3) (RATIONAL) REFLECTION

But in fact these claims are not, on their own, inconsistent. There is a wide class of non-partitioning experiments in which the conditionalizer may satisfy REFLECTION. (And likewise for RAT-REF.) Suppose you will conduct the non-partitioning experiment $\mathcal{E} = \{1, 2, 3, \neg 1, \neg 2, \neg 3, \top\}$ with the prior credence distribution shown in figure 3a. An exercise for the reader: pick any proposition ϕ (any disjunction of non-zero cells) from the 3×7 grid in figure 3a, and then calculate both $C(\phi)$ and $\sum_{e \in \mathcal{E}} C(\phi \mid e) \cdot C(\mathbf{U}e)$. You will find that they are equal, no matter which of the $2^{12} = 4{,}096$ possible propositions you pick.

The additional assumption needed to get a contradiction out of (G1), (G2), and (G3) is that you may conduct an experiment like the one shown in figure 4. An externalist could deny that you may find yourself in experiments like these, but this would strain credibility. Suppose you find yourself in an experiment in which you might learn anything about the location of the ball, and then, a trustworthy confidant informs you that Bonnie definitely won't show you the ball, and she definitely won't reveal what's under cup 2. You should then be certain that the strongest thing you'll learn about the position of the ball is either $\neg 1$ or $\neg 3$; and it's unclear how this new information you've acquired would make any difference with respect to whether you will additionally gain introspective evidence about how your credences have changed.

So the externalist faces a choice: they can either deny CONDI or they can deny both REFLECTION and RAT-REF. ELGA (2007, 2013) affords an argument for rejecting RAT-REF:[18] As I made the case for cutting to the chase, I said: prior to conducting the experiment, you are certain that your future rational self will know all that you know and more besides. It was on this basis that I recommended deferring to their opinion about whether 2. Perhaps this claim was in error. Perhaps there is something you now know that your future rational self does not. For you now know that your future rational self *is rational.* In particular, you know that they

17 I believe that the first to explicitly note this inconsistency was HILD (1998a,b).

18 ELGA only argues for a *synchronic* principle saying how your *current* credence that ϕ should relate to your *current* credences about which credence that ϕ is rational. Here, I generalize his discussion—perhaps in ways he would not endorse.

have conditioned on their total evidence. However, if your total evidence ends up being $\neg 1$, then your rational future self will give positive credence to their total evidence being $\neg 3$. And, if their total evidence is $\neg 3$, then they are irrationally confident in 3. Similarly, if your total evidence ends up being $\neg 3$, then your rational future self will give non-zero credence to their total evidence being $\neg 1$. And, if their total evidence is $\neg 1$, then they are irrationally confident in 1. So your future rational self will not be certain that they are rational, though you are now certain that they will be. In general, you should not defer to agents when you are certain of matters about which they are ignorant; rather, you should defer to them only *after* apprising them of the information you have that they lack. So, you should only defer to the opinion of your future rational self *after* informing them of what their total evidence is. Once they have this extra information, both of your future rational selves will give the proposition 2 credence 1/3, since $C(2 \mid \neg 1 \cap \mathbb{T}\neg 1) = C(2 \mid \neg 3 \cap \mathbb{T}\neg 3) = 1/3$.[19] In general, the externalist should defer to their rational post-experimental self in the manner prescribed by 'NEW RATIONAL REFLECTION'.[20]

> NEW RATIONAL REFLECTION
> Your pre-experimental credence that ϕ should be equal to your expectation of your rational post-experimental credence that ϕ, *once your rational post-experimental self has been told what its total evidence is.*
>
> (NEW RAT-REF) $$C(\phi) \stackrel{!}{=} \sum_{e \in \mathcal{E}} C_e(\phi \mid \mathbb{T}e) \cdot C(\mathbb{T}e)$$

Holding fixed the factivity of evidence, NEW RAT-REF follows from CONDI. Since EXTERNALISM and CONDI are consistent, EXTERNALISM, CONDI, and NEW RAT-REF are consistent as well.

This diagnosis of the conflict between (G1), (G2), and (G3) is clever but not ultimately persuasive. The foregoing considerations do nothing to blunt the argument for cutting to the chase. Even supposing that, before carrying out the experiment $\mathcal{E} = \{\neg 1, \neg 3\}$, you know something that your future rational self does not—*viz*, what their total evidence is—this knowledge of yours is *irrelevant* with respect to the question of what degree of belief in the proposition 2 is rational. It's true that, if your evidence is $\neg 1$, then your post-experiment self will think they may be overly confident in 3, and if your evidence is $\neg 3$, then they'll think they may be overly confident in 1. However, in either case, they will be certain that their credence that 2 is rational. For they can, post-experiment, run

[19] See HALL (1994)'s distinction between a *database expert* and an *analyst expert* and ELGA's distinction between an *expert* and a *guru*.

[20] Again, the principle ELGA calls 'new rational reflection' is synchronic.

precisely the same argument that you are able to run pre-experiment. They can say to themselves: "Either my total evidence was $\neg 1$ or it was $\neg 3$. If it was $\neg 1$, then it's rational for me to have credence 1/2 that 2. If it was $\neg 3$, then it's rational for me to have credence 1/2 that 2. So, either way, it's rational for me to have credence 1/2 that 2." Post-experiment, no matter what you learn, you will be rationally certain that 1/2 is the rational credence to have in the proposition 2. It's hard to see why the fact that your future rational self is uncertain about whether their credence in some other proposition is rational gives any reason to not adopt their credence in the proposition 2, which is certainly rational.

In the second place, it's simply not the case that, pre-experiment, you know something that your rational post-experimental self does not. It's true that your future rational self does not have the evidence of what their total evidence is, but neither does your pre-experimental self. Likewise, it's true that your future rational self doesn't know whether a post-experimental credence of 1/2 in 1 is rational; but neither does your pre-experimental self. Your post-experimental rational self has all the evidence your pre-experimental self has, and more besides. And they share your precise epistemic standards. So there is no reason why you should not regard their opinion as better informed than your own. Assuming CONDI, they are certain to have a rational degree of belief of 1/2 that 2. So, assuming CONDI, you too should have a credence of 1/2 that 2.

That's not an endorsement of REFLECTION or RAT-REF. Nor is it a criticism of NEW RAT-REF. It is simply an endorsement of REFLECTION and RAT-REF's advice to the conditionalizer conducting the experiment $\mathcal{E} = \{\neg 1, \neg 3\}$: don't wait—cut to the chase. The advice is sound, but it is inconsistent with CONDI. If the externalist is or should be a conditionalizer, they shouldn't be. So they shouldn't be. For all we've said so far, perhaps the externalist should *also* reject REFLECTION, RAT-REF, and/or NEW RAT-REF. But, at a minimum, they should reject CONDI. (To lay my cards on the table: the update I offer the externalist in §4 entails both REFLECTION and NEW RAT-REF.)

3 Reflection and Biased Inquiry

Externalists face a choice between the principles of REFLECTION and CONDITIONALIZATION. They cannot plausibly endorse both. I argued that, whatever externalists should think of REFLECTION generally, they should accept its recommendations in one particular experiment; and, in that experiment, its recommendations conflict with CONDITIONALIZATION. So externalists should reject CONDITIONALIZATION. There is additionally reason to accept the principle of REFLECTION in full generality.

SALOW (forthcoming) teaches that REFLECTION is more than a principle of expert deference. Whether or not your post-experimental self is worthy of epistemic deference, the principle of REFLECTION has important work to do in preventing rational agents from engaging in deliberate self-delusion. If your update strategy violates REFLECTION, you will expect to raise your credence in some proposition, irrespective of whether or not that proposition is true.

To borrow SALOW's example: let 'p' be the proposition that you are popular (or that you're not—whichever you'd prefer to believe). Suppose that your rational degree of belief in p is 1/3 (though the precise value won't matter). Suppose that it's possible for you to conduct the experiment $\mathcal{E} = \{\neg 1, \neg 3\}$ from figure 4, and assume CONDI. Then, here's a recipe for raising your credence that p no matter what. First, tell a confidant who knows the truth about p to place the ball under cup 2 iff p. If $\neg p$, then they should flip a coin to decide between cup 1 and 3. Tell them to reveal to you an empty cup, but not under any circumstance to reveal what's under cup 2. CONDI counsels to raise your credence that 2 from 1/3 to 1/2, no matter what. Since you are certain that $p \longleftrightarrow 2$, your credence that p will likewise rise from 1/3 to 1/2. And there's no reason this experiment need be conducted only once. Run through the whole exercise again, and plan to raise your credence that p to 2/3, no matter what; and—why not?—again, raising it to 4/5 no matter what; and again, raising it to 8/9; and so on and so forth. If it is within your power to design experiments like $\mathcal{E} = \{\neg 1, \neg 3\}$, and if it is rational to strategize with CONDI, then it is rational to plan to become as confident in the proposition that you're popular as you wish. Assuming it is rational to follow through on a rationally-formed plan, your future rational self will be as confident that p as you wish them to be.

Let us be frank: this is not rational inquiry. This is self-delusion, no more. And no sensible epistemology will deem rational the person who plans to walk away from this series of experiments nearly certain that they are popular. This is non-negotiable. So let us lay it down as a principle.

> NO SELF-DELUSION
> A rational agent may not design an experiment and strategize to become more confident in some proposition, no matter the experiment's outcome.

Granting that experiments like $\{\neg 1, \neg 3\}$ from figure 4 are possible, the following are inconsistent:

H1) EXTERNALISM

H2) CONDITIONALIZATION

H3) NO SELF-DELUSION

NO SELF-DELUSION is non-negotiable. A rational agent cannot structure their inquiry so as to become arbitrarily confident in a proposition, no matter what. SALOW concludes that EXTERNALISM is false. Perhaps that is the correct lesson to draw. However, I believe that a plausible version of EXTERNALISM is left standing. This is a version of EXTERNALISM which accepts NO SELF-DELUSION by denying CONDITIONALIZATION. In the following section, I will provide the externalist with an alternative to CONDITIONALIZATION. This alternative will always abide by NO SELF-DELUSION.

NO SELF-DELUSION prohibits an extreme variety of biased inquiry—inquiry which is *guaranteed* to leave you more confident of some proposition. The reasons we have to prohibit this kind of biased inquiry carry over to inquiries which we merely *expect* to leave us more confident in some proposition. Say that an update strategy is biased in favor of a proposition ϕ iff, when enacting that strategy, your expectation of your post-experimental credence that ϕ is greater than your pre-experimental credence that ϕ. Similarly, an update strategy is biased against ϕ iff your expectation of your post-experimental credence that ϕ is less than your pre-experimental credence that ϕ. Thus, an update strategy is unbiased iff, for all ϕ,

$$C(\phi) = \sum_{e \in \mathcal{E}} C_e(\phi) \cdot C(\mathbf{U}e)$$

Given this understanding of when an update strategy is biased, SALOW endorses:

> NO BIASED INQUIRY
> A rational agent will not have a biased update strategy.

SALOW's insight is that the principle of REFLECTION is equivalent to NO BIASED INQUIRY. To strategize to update in defiance of REFLECTION is to bias your inquiry. Biasing your inquiry is irrational, so REFLECTION is rationally required. The externalist update I will propose in the following section will entail the principle of REFLECTION, and so will never permit biased inquiry.

4 Updating for Externalists

I offer a new update strategy to the externalist; but mine is not the only externalist update on the market. In my younger and more vulnerable years,[21] I offered an update custom tailored to handle cases in which the Brouwer principle is violated—cases in which e is false, though your evidence doesn't rule out that you have the evidence e. I understood these as cases in which your evidence is theory-dependent. There are two hypotheses: that Sabeen has slipped you a psychotropic drug, d, and that she has not, $\neg d$. The drug renders you incapable of properly categorizing flavor experiences. Without the drug, you are able to recognize how things taste to you. With the drug, your beliefs about how things taste to you correlate not at all with the way they actually taste to you. You bite into the pear. In this case, I said: what your evidence is depends upon which background theory

[21] GALLOW (2014)

is true. If you are drug-free, then your evidence is that the pear tastes sweet, s. If drugged, you have no evidence at all, $\top$. I suggested representing the input to your update with the set of ordered pairs, $\{<\neg d, s>, <d, \top>\}$. More generally, in cases of theory-dependent evidence, you will have an input $\{<t_i, e_i>\}_i$, with the interpretation that for each i, if t_i is true, then your evidence is e_i. Then, I endorsed HOLISTIC CONDITIONALIZATION.

> HOLISTIC CONDITIONALIZATION
> If C is your pre-experimental credence function and $\{<t_i, e_i>\}_i$ is the input acquired in the experiment, then your rational post-experimental credence function, C^+, is such that, for every proposition ϕ,
>
> (HCONDI) $$C^+(\phi) = \sum_i C(\phi \mid t_i \cap e_i) \cdot C(t_i)$$

We have up to this point agreed to use 'evidence' to mean 'certainty evidence'. So, on our use of 'evidence', 'your evidence is e' entails that certainty in e has been rationalized. HCONDI tells you to be less than certain in your evidence. So, what I formerly called 'evidence' is not what we have decided to call 'evidence' here. Still, we can translate my former claims into our current idiom. When I formerly said 's is your total evidence if the background theory $\neg d$ is true and $\top$ is your total evidence if d', I added that $\neg d \rightarrow s$ is the strongest proposition about which certainty is rationalized. And when I said '$\neg s$ is your total evidence if $\neg d$ and $\top$ is your total evidence if d', I added that $\neg d \rightarrow \neg s$ is your the strongest proposition about which certainty is rationalized. So, in this 'theory-dependent' experiment, your experience will either rationalize certainty in $\neg d \rightarrow s$ or $\neg d \rightarrow \neg s$, and your total *certainty* evidence will either be $\neg d \rightarrow s$ or $\neg d \rightarrow \neg s$. So, in this case, your experiment is not a partition. So HCONDI is an externalist update.

HCONDI correctly handles some externalist experiments, but it cannot handle all. Experiments like taking a glance at Williamson's unmarked clock do not plausibly involve any theory-dependent evidence. In such cases, HCONDI reduces to CONDI. So, HCONDI can lead to rational self-delusion in these cases. HCONDI has other problems as well. For instance, it holds fixed your credence in the various background theories, t_i. But suppose you know in advance that the chance of the pear tasting sour is one in a million, while the chance of Sabeen slipping you the drug is one half. Then you should plan to be very confident you've been slipped the drug if the pear tastes sour.[22]

HILD (1998a,b) and SCHOENFIELD (forthcoming) both endorse an update

[22] I recognized this problem at the time, and suggested a rule for updating your credence in background theories, but the suggestion is overly complicated and insufficiently motivated, and I no longer endorse it.

strategy which we may call 'EVIDENTIAL CONDITIONALIZATION'.[23] According to this strategy, you should update by conditioning, not on your total evidence, but rather on the proposition *that it is* your total evidence.

> EVIDENTIAL CONDITIONALIZATION
> If you conduct the experiment $\mathcal{E}$, with the pre-experimental credence function C, then, for each $e \in \mathcal{E}$, your strategy for responding to total evidence e, C_e, should be to condition C on the proposition that e is your total evidence. That is, for all propositions ϕ,
>
> (EVCONDI) $$C_e(\phi) \stackrel{!}{=} C(\phi \mid \mathbf{T}e)$$
>
> where '$\mathbf{T}e$' is the proposition that e is your total evidence.

HILD and SCHOENFIELD's use of 'evidence' also differs from our own. They each think that e can be your total evidence even when the proposition that e is your total evidence is stronger than e itself. On our use of 'total evidence', if $\mathbb{T}e$ is true, and $\mathbb{T}e$ is stronger than e, then—*by definition!*—it cannot be rational to become certain that $\mathbb{T}e$. Yet EVCONDI tells you to become certain that e is your total evidence even then. So, in EVCONDI, 'total evidence' cannot mean 'strongest proposition about which certainty has been rationalized'. When HILD or SCHOENFIELD says 'your total evidence is e', I will write '$\mathbf{T}e$', to distinguish their terminology from our own. As HILD and SCHOENFIELD use 'total evidence', EVCONDI is an update strategy for externalists. However, as we have chosen to use the term, EVCONDI entails both INTERNALISM and CONDITIONALIZATION.

Proof. Your experiment contains all propositions which may be your total evidence. If HILD and SCHOENFIELD say $\{e_1, e_2, \ldots, e_N\}$ is your experiment, then $\{\mathbf{T}e_1, \mathbf{T}e_2, \ldots, \mathbf{T}e_N\}$ forms a partition. If HILD and SCHOENFIELD call e_i your total evidence, then EVCONDI says it is rational for you to become certain that $\mathbf{T}e_i$, and nothing stronger. So we call $\mathbf{T}e_i$ your total evidence. So we say that $\{\mathbf{T}e_1, \mathbf{T}e_2, \ldots, \mathbf{T}e_N\}$ is your experiment. So what we call your experiment forms a partition, which is equivalent to INTERNALISM. EVCONDI says to conditionalize on what we call your total evidence. So EVCONDI entails CONDITIONALIZATION. □

So whatever merit EVCONDI may otherwise have, it is not an externalist update, as we are using that term.

Think of it like this: experience may teach that e.[24] If so, say that e is *experiential* evidence for you. If experience teaches that e and no more, then say

[23] HILD calls the rule 'Auto-epistemic conditionalization', and SCHOENFIELD calls it 'conditionalization*'.

[24] 'Experience teaches that' is a broad and ecumenical notion. I assume very little about it. I do not even assume that it is factive. Perhaps experience could teach that e even when e is false.

that e is your *total* experiential evidence, and write '$\mathbf{T}e$'. If experience may teach each $e_i \in \{e_1, e_2, \ldots, e_N\}$ (and no more), then say that $\{e_1, e_2, \ldots, e_N\}$ is your experiential experiment. If $\{e_1, e_2, \ldots, e_N\}$ is your experiential experiment, then the possibilities are partitioned by $\{\mathbf{T}e_1, \mathbf{T}e_2, \ldots, \mathbf{T}e_N\}$. HILD and SCHOENFIELD's 'evidence' is our 'experiential evidence'. Their 'total evidence' is our 'total experiential evidence'.

To avoid confusion: if e is the strongest proposition about which certainty has been rationalized, then let us call e your 'total certainty evidence'. Likewise, call the set of propositions which may be your total certainty evidence your 'certainty experiment'. If rational certainty supervenes upon the lessons of experience, then *experiential evidence* and *experiential experiment* are the strictly more general notions. For two pieces of experiential evidence could rationalize certainty in one and the same proposition, thereby corresponding to the same certainty evidence. By supervenience, the same may not be said in reverse; the same experiential evidence must correspond to the same certainty evidence.

We have been considering updates which specify a post-experimental credence function for each proposition in your certainty experiment (call them 'certainty updates'). If experiential experiments are strictly more general than certainty experiments, then we should instead consider strategies for experiential experiments. Call such a strategy an 'experiential update'. An experiential update will specify a post-experimental credence function for each proposition in your experiential experiment. HILD and SCHOENFIELD's EVCONDI is an experiential update, and not a certainty update.

Suppose you have your experiential update. In response to an experience teaching that e, you plan to adopt a new credence function—call it 'C_e' and let '$\mathbf{U}e$' be the proposition that you have updated to C_e. Take a simple case: experience will either teach e_1 (and no more) or e_2 (and no more). The propositions e_1 and e_2 are consistent, so $\{e_1, e_2\}$ does not partition the possibilities. Even so, $\{\mathbf{T}e_1, \mathbf{T}e_2\}$ is a partition. Suppose you are immodest—for each i, you are certain that $\mathbf{U}e_i \leftrightarrow \mathbf{T}e_i$.[25] Then, the partitions $\{\mathbf{U}e_1, \mathbf{U}e_2\}$ and $\{\mathbf{T}e_1, \mathbf{T}e_2\}$ align; they draw the same distinction. In this case, EVCONDI says that your response to an experience teaching e_i should be certainty in $\mathbf{T}e_i$. So your total certainty evidence will be $\mathbf{T}e_1$ iff $\mathbf{T}e_1$ is true, and $\mathbf{T}e_2$ iff $\mathbf{T}e_2$ is true. So the partitions $\{\mathbb{T}\mathbf{T}e_1, \mathbb{T}\mathbf{T}e_2\}$ and $\{\mathbf{T}e_1, \mathbf{T}e_2\}$ will likewise align. Your certainty experiment will form a partition, even though your experiential experiment does not. (See figure 5.)

HILD and SCHOENFIELD think your experiential experiment need not form a partition, though they insist that your certainty experiment always will. This

[25] Before, I said you were immodest iff you were certain that $\mathbf{U}e \leftrightarrow \mathbb{T}e$, for each e in your certainty experiment. Now, I say you are immodest iff you are certain that $\mathbf{U}e \leftrightarrow \mathbf{T}e$, for each e in your experiential experiment.

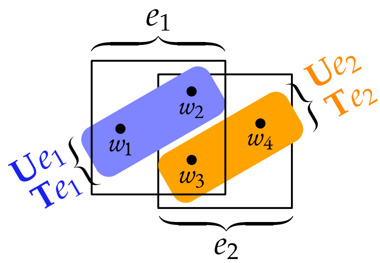

FIGURE 5: $e_1 = \{w_1, w_2, w_3\}$ and $e_2 = \{w_2, w_3, w_4\}$. In w_1 and w_2, your total experiential evidence is e_1. In w_3 and w_4, your total experiential evidence is e_2. In both w_1 and w_2, you correctly update to $C_{e_1}, \mathbf{U}e_1$, and in both w_3 and w_4 you correctly update to $C_{e_2}, \mathbf{U}e_2$. EVCONDI says that, in w_1 and w_2, your total certainty evidence is $\mathbf{T}e_1$; while, in w_3 and w_4, your total certainty evidence is $\mathbf{T}e_2$. So your certainty experiment forms a partition.

is a thesis worth calling 'externalism', though it is not the thesis we have called 'EXTERNALISM'.[26] Call the former thesis 'lowercase externalism'—the latter, 'uppercase externalism'. EVCONDI is a rule for lowercase externalists. Though it permits non-partitional experiential experiments, it does not allow non-partitional certainty experiments.

SCHOENFIELD argues for EVCONDI as follows. Represent an experiential update strategy, σ, with a function from propositions of the form $\mathbf{T}e$ to post-experimental credences.[27] If $\sigma(\mathbf{T}e) = C$, then C is the post-experimental credence function σ says to adopt if your total experiential evidence is e. Choose some measure of the *accuracy* of a credence function at a world, $\mathscr{A}(C, w)$—but choose so that $\mathscr{A}$ is strictly proper.[28] An update strategy is good to the extent that you expect its outputs to be accurate.[29] And the strategy which maximizes expected accuracy is EVCONDI.[30] You should select an update strategy which maximizes expected accuracy. So you should strategize to update with EVCONDI.[31]

I do not disagree with HILD or SCHOENFIELD. Their focus is narrower than mine. Their rule is for lowercase externalists who are not uppercase externalists. Neither EVCONDI nor SCHOENFIELD's accuracy argument in its favor apply in con-

---

[26] This externalist thesis is contested—LEWIS (1996, 1999) holds that your experiential experiment will always be a partition. (I assume that, for LEWIS, experience teaches that e iff e is describes your experience in full detail.)

[27] Henceforth, when I say 'update' or 'update strategy', I will mean *experiential* update strategy.

[28] $\mathscr{A}$ is strictly proper iff every probability function expects itself to have a strictly higher $\mathscr{A}$-value than any other credence function. See ODDIE (1997), GIBBARD (2008), PREDD et al. (2009), JOYCE (2009), and PETTIGREW (2012) for more on strict propriety.

[29] That is: the goodness of an update strategy σ is given by $\sum_e \sum_{w \in \mathbf{T}e} C(w) \cdot \mathscr{A}(\sigma(\mathbf{T}e), w)$.

[30] This follows from Corollary 2 in GREAVES & WALLACE (2006).

[31] HILD (1998b) offers a diachronic Dutch-book argument for EVCONDI. Like SCHOENFIELD's accuracy argument, it presupposes that you are immodest.

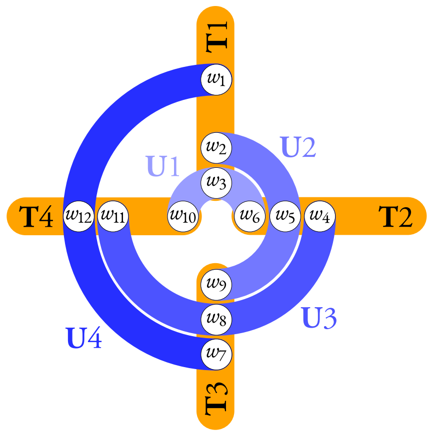

Figure 6: w_4, w_5, and w_6 are each possibilities at which experience teaches that the clock hand is at position 2, **T**2. At w_5, you correctly update on the information that 2, **U**2. While, at w_4, you incorrectly update on the information that 3, **U**3. And, at w_6, you incorrectly update on the information that 1, **U**1.

ditions of modesty.

Consider a simplified version of Williamson's clock. You know in advance that the clock hand will point at one of four positions. Call them '1', '2', '3', and '4'. (This example will simplify past the point of psychological plausibility, but what we learn from the simple case will carry over to the more complicated cases, so it's a harmless abstraction.) If the clock hand is at position n, then experience will teach that it is at position n. However, you are less than fully confident in your ability to correctly learn from experience. You modestly foresee possibilities of error. In particular, if the clock hand is at position 2, you may correctly update on the information that 2, but you may also incorrectly update on the information that 1 or 3 instead. And, *mutatis mutandis*, the same is true if the clock hand is at position 1, 3, or 4. In each case, while you may learn the correct lesson from experience, you modestly foresee the possibility that you err, either clockwise or counterclockwise by a single position. (See figure 6.)

A parable: you stand in a room with a television screen and four buttons, arranged in a circle. Your credences are displayed outside of the room; since there are no windows, you cannot see them. The screen will show a number, 1 through 4. If you see the number n, you will attempt to push button n. Each button triggers a different update to your credences outside. Pressing button 1 will trigger the first update, pressing button 2 will trigger the second, and similarly for buttons 3 and 4. The buttons are close together and your hands shake, so sometimes, when you try to push button 1, you end up pushing button 4 or 2 instead. Sometimes, when you try to push button 2, you end up pushing button 1 or 3 instead. And similarly for buttons 3 and 4. When you press the wrong button, you do not learn that you have erred. So, once a button is pressed, you are not in a position to know what your credences are. The television screen is your experience. The buttons are your plans for update. You are modest; your plans may go awry, in which case your credences will not change as planned. You lack introspective access, wherefore these errors may not be detected.

Which update should each button trigger? Answer #1: button n should trigger an update to certainty that the television displays number n, $\mathbf{T}n$. So long as you press the right button (so long as you respond rationally) you will be certain of a truth. Of course, if you press the wrong button (if you respond *ir*rationally) then you will be certain of a falsehood. But this shows only that you should not press the wrong button (you should not be irrational). Answer #2: button n should trigger an update to certainty that update n has taken place, $\mathbf{U}n$. You may press the wrong button, but even so, your credences will end up being accurate. The screen shows 2 and by accident you press button 1. You end up certain that update 1 was triggered, contrary to plan. But, lo and behold, update 1 *was* triggered—the error has undone itself! Answer #3: Answers #1 and #2 are both right—button n should trigger an update to certainty in *both* $\mathbf{T}n$ *and* $\mathbf{U}n$. That's because you should be certain, before pushing, that $\mathbf{U}n \leftrightarrow \mathbf{T}n$. That is: you should be immodest. Updating on something other than what experience has taught is irrational. Rationality requires certainty that you are and will remain rational. So rationality requires immodesty. Answer #4: button 1 should only trigger an update to certainty that the screen did not display 3. Similarly, button 2 should only trigger an update to certainty that the screen did not display 4. And likewise for buttons 3 and 4. Suppose the television screen shows the number n; in response, you attempt to push button n. Your only reasons to think you succeeded in pushing button n are statistical, so it is irrational to be *certain* that $\mathbf{U}n$. And if button n were to mandate certainty that $\mathbf{T}n$, you would open yourself up to the possibility of being certain of a falsehood, should you press the wrong button. Opening yourself up to this possibility is irrational. So certainty in either $\mathbf{T}n$ or $\mathbf{U}n$ is irrational.

To each of these answers corresponds an accuracy-maximization argument like Schoenfield's. In this framework, which answer we endorse depends upon which kinds of update strategies we take to be the objects of choice. Suppose you must choose between functions, $\sigma_{\mathbf{T}}$, from the partition $\{\mathbf{T}1, \mathbf{T}2, \mathbf{T}3, \mathbf{T}4\}$ to post-experimental credence functions. Then, the choice of $\sigma_{\mathbf{T}}$ which maximizes expected accuracy will be such that $\sigma_{\mathbf{T}}(\mathbf{T}n) = C(- \mid \mathbf{T}n)$, for each n (this is just evcondi). Answer #1 is vindicated. Suppose, on the other hand, that you must choose between functions, $\sigma_{\mathbf{U}}$, from the partition $\{\mathbf{U}1, \mathbf{U}2, \mathbf{U}3, \mathbf{U}4\}$ to post-experimental credence functions. Then, the available strategy which maximizes expected accuracy will be such that $\sigma_{\mathbf{U}}(\mathbf{U}n) = C(- \mid \mathbf{U}n)$, for each n. Answer #2 is vindicated. Answer #3 sees the two choices above as equivalent. For answer #3 requires you to be immodestly certain that $\mathbf{U}n \leftrightarrow \mathbf{T}n$, for each n. So it requires the partitions $\{\mathbf{T}1, \mathbf{T}2, \mathbf{T}3, \mathbf{T}4\}$ and $\{\mathbf{U}1, \mathbf{U}2, \mathbf{U}3, \mathbf{U}4\}$ to align.

The uppercase externalist should be unhappy with each of the vindications above. They should say: an update strategy is available as an object of choice iff 1) its inputs are your total experiential evidence; and 2) you are in a position to control which of its outputs obtain. In the vindication of answer #2, we took the inputs to your strategy to be the button pressed, $\mathbf{U}n$, and not the total evidence delivered by the television screen. So this strategy is not available as an object of choice for you. In the vindication of answer #1, we took the inputs to your strategy to be the number displayed on the screen, $\mathbf{T}n$, but we took its outputs to be post-experimental credence functions. Because your hands shake, you are not in a position to control which post-experimental credence you adopt if $\mathbf{T}n$. So the strategy to be certain of $\mathbf{T}n$ iff the screen shows n is not an object of choice for you, either. (For the same reason, the strategy of answer #3 is not available as an object of choice.)

Consider figure 6. At w_4, w_5, and w_6, the television screen displays 2. You must choose what to do in this contingency, so you should have the same strategy at each of these possibilities. But you press different buttons in each of these possibilities. In w_5, you successfully press button 2. But in w_4, you incorrectly press button 1. And in w_6, you incorrectly press button 3. If the screen displays 2, whether you press button 2 is not under your control. Nevertheless, something still is under your control: it is under your control which button to *attempt* to press, and therefore, it us under your control with what probability you will end up pressing button 1, 2, or 3. When you choose to attempt to press button 2 in w_4, w_5, and w_6, what you choose is to update to C_1 with probability $C(\mathbf{U}1 \mid \mathbf{T}2)$, to update to C_2 with probability $C(\mathbf{U}2 \mid \mathbf{T}2)$, and to update to C_3 with probability $C(\mathbf{U}3 \mid \mathbf{T}2)$. Though no *pure* strategy is available to you, this *mixed* strategy is.

A pure strategy, σ, is a function from a partition to post-experimental credence functions. A *mixed* strategy, μ, is a function from a partition to probability distributions over post-experimental credence functions. In our simplified version of Williamson's clock, a mixed strategy is available for you as an object of choice just in case its inputs come from $\{\mathbf{T}1, \mathbf{T}2, \mathbf{T}3, \mathbf{T}4\}$ and there are credence functions C_1, C_2, C_3, and C_4 such that:

$$
\begin{aligned}
\mu(\mathbf{T}1) &= C(\mathbf{U}1 \mid \mathbf{T}1) \cdot C_1 + C(\mathbf{U}2 \mid \mathbf{T}1) \cdot C_2 + C(\mathbf{U}4 \mid \mathbf{T}1) \cdot C_4 \\
\mu(\mathbf{T}2) &= C(\mathbf{U}1 \mid \mathbf{T}2) \cdot C_1 + C(\mathbf{U}2 \mid \mathbf{T}2) \cdot C_2 + C(\mathbf{U}3 \mid \mathbf{T}2) \cdot C_3 \\
\mu(\mathbf{T}3) &= C(\mathbf{U}2 \mid \mathbf{T}3) \cdot C_2 + C(\mathbf{U}3 \mid \mathbf{T}3) \cdot C_3 + C(\mathbf{U}4 \mid \mathbf{T}3) \cdot C_4 \\
\mu(\mathbf{T}4) &= C(\mathbf{U}1 \mid \mathbf{T}4) \cdot C_1 + C(\mathbf{U}3 \mid \mathbf{T}4) \cdot C_3 + C(\mathbf{U}4 \mid \mathbf{T}4) \cdot C_4
\end{aligned}
$$

In general, the externalist should say: when conducting the experiential experiment $\mathcal{E}$, a mixed update strategy μ is *available* iff it takes as inputs propositions of the form $\mathbf{T}e$ and, for each $e \in \mathcal{E}$,

$$\mu(\mathbf{T}e) = \sum_{f \in \mathcal{E}} C(\mathbf{U}f \mid \mathbf{T}e) \cdot C_f$$

When selecting your update strategy, you are not choosing the probabilities $C(\mathbf{U}f \mid \mathbf{T}e)$. You are only choosing the post-experimental credence functions C_f, for each $f \in \mathcal{E}$. Which choice maximizes expected accuracy?

To answer this question, we must first say something about how to measure the expected accuracy of a mixed update strategy. If $\mathscr{A}(C', w)$ is our measure of the accuracy of a credence function C' at world w, then the expected accuracy of a *pure* update strategy σ is $\sum_e \sum_{w \in \mathbf{T}e} C(w) \cdot \mathscr{A}(\sigma(\mathbf{T}e), w)$. If μ is a mixed update strategy, then we cannot say for sure which post-experimental credence function it delivers. So we cannot say for sure what the accuracy of μ is at any given world.[32] Still, we can say what its *expected* accuracy is at each world. If μ is available, then its expected accuracy at world $w \in \mathbf{T}e$ will be $\sum_{f \in \mathcal{E}} C(\mathbf{U}f \mid \mathbf{T}e) \cdot \mathscr{A}(C_f, w)$. Then, the expected accuracy of μ will be

$$\sum_{e \in \mathcal{E}} \sum_{w \in \mathbf{T}e} C(w) \cdot \sum_{f \in \mathcal{E}} C(\mathbf{U}f \mid \mathbf{T}e) \cdot \mathscr{A}(C_f, w)$$

Suppose you begin with the pre-experimental credence function from figure 7a. You think the clock hand is equally likely to be at any of the four positions. If experience teaches that the clock hand is at position 2, then you think it 80% likely that you'll correctly adopt C_2, though you foresee some possibility of error, so you save 20% of your credence for incorrectly adopting C_1 or C_3 instead, each with equal probability. And the same is true not just for position 2, but for positions 1, 3, and 4 as well, *mutatis mutandis*. Suppose you wish to select an available mixed update strategy which maximizes expected accuracy, and you measure accuracy with a strictly proper, additive, and extensional measure.[33] Then, the optimal choice of C_2 is shown in figure 7b. After updating on the clock hand's being at 2, you will think that it is most likely at position 2 (80%), though you'll save some credence for it being at 1 or 3 instead (10% each). While you'll think that your total experiential evidence was most likely 2 (80%), you will also think that

[32] This point is subtle. Consider figure 6. We *do* know that, *e.g.*, at w_5, adopting the mixed strategy μ will result in you adopting the credence function C_2. Even so, asking about the expected accuracy of adopting the strategy μ at w_5 is not the same as asking about the expected accuracy of adopting a *pure* strategy which maps w_5 to C_2. The latter asks about a strategy which is *guaranteed* to result in the credence C_2; whereas the former asks about a strategy which has some probability of resulting in C_1 or C_3 instead.

[33] I say an accuracy measure $\mathscr{A}$ is 'additive' iff the accuracy of a credence function C in world w, $\mathscr{A}(C, w)$, is the sum of the accuracy of C's credence in ϕ at w, $\mathscr{A}(C(\phi), \phi, w)$, for each proposition ϕ. I say that $\mathscr{A}$ is *extensional* iff the accuracy of a credence $C(\phi)$ in the proposition ϕ at a world w depends only on $C(\phi)$ and the truth-value of ϕ at w.

	$1 \cap \mathbf{T}1$	$2 \cap \mathbf{T}2$	$3 \cap \mathbf{T}3$	$4 \cap \mathbf{T}4$
U1	8/40	1/40	0	1/40
U2	1/40	8/40	1/40	0
U3	0	1/40	8/40	1/40
U4	1/40	0	1/40	8/40
	1/4	1/4	1/4	1/4

(a)

	$1 \cap \mathbf{T}1$	$2 \cap \mathbf{T}2$	$3 \cap \mathbf{T}3$	$4 \cap \mathbf{T}4$
U1	8/100	8/100	0	0
U2	1/100	64/100	1/100	0
U3	0	8/100	8/100	0
U4	1/100	0	1/100	0
	1/10	8/10	1/10	0

(b)

Figure 7: Given the prior credal state in figure 7a, the result of updating on 2 with excondi is shown in figure 7b.

it could have been either 1 or 3 (10% each). And you'll think that, most likely, you have updated on 2 (66%), though you may have updated on 1 or 3 instead (16% each), and you'll even put aside some credence (2%) for the possibility that you've updated on 4. Note that, though your total *experiential* evidence is that the clock hand is at 2, **T**2, your total *certainty* evidence is merely that it is not at 4, $\mathbb{T}\neg 4$.

More generally: suppose you conduct the experiential experiment $\mathcal{E}$, which may or may not be a partition. Suppose you measure accuracy with a strictly proper, additive, and extensional measure, and you wish to select an available mixed strategy which maximizes expected accuracy. Then, you should select the strategy I will call 'excondi'.

externalist conditionalization

If you conduct the experiential experiment $\mathcal{E}$ with the pre-experimental credence function C, then, for each $e \in \mathcal{E}$, your strategy for responding to the total experiential evidence e, C_e, should be, for each $f \in \mathcal{E}$, to change your credence that experience has taught that f, $\mathbf{T}f$, to your pre-experimental credence in $\mathbf{T}f$ given $\mathbf{U}e$, and leave alone your credence in all propositions conditional on $\mathbf{T}f$. That is, for each ϕ,

(excondi) $$C_e(\phi) \overset{!}{=} \sum_{f \in \mathcal{E}} C(\phi \mid \mathbf{T}f) \cdot C(\mathbf{T}f \mid \mathbf{U}e)$$

(**Proposition 1** in the appendix shows that excondi is the available mixed strategy which maximizes expected accuracy, given any strictly proper, additive, and extensional measure of accuracy.)

Note that EXCONDI entails REFLECTION.

Proof.

$$\begin{aligned}
\sum_{e \in \mathcal{E}} C_e(\phi) \cdot C(\mathbf{U}e) &\overset{!}{=} \sum_{e \in \mathcal{E}} \sum_{f \in \mathcal{E}} C(\phi \mid \mathbf{T}f) \cdot C(\mathbf{T}f \mid \mathbf{U}e) \cdot C(\mathbf{U}e) \\
&= \sum_{f \in \mathcal{E}} C(\phi \mid \mathbf{T}f) \cdot \sum_{e \in \mathcal{E}} C(\mathbf{T}f \mid \mathbf{U}e) \cdot C(\mathbf{U}e) \\
&= \sum_{f \in \mathcal{E}} C(\phi \mid \mathbf{T}f) \cdot C(\mathbf{T}f) \\
&= C(\phi)
\end{aligned}$$

□

So EXCONDI will never permit self-delusion; it will always abide by NO BIASED INQUIRY. Note also that EXCONDI entails ELGA (2013)'s 'new rational reflection' principle.[34]

Proof. First, note that, for any ϕ and any $e \in \mathcal{E}$, $C_e(\phi \mid \mathbf{T}e) = C(\phi \mid \mathbf{T}e)$.

$$C_e(\phi \mid \mathbf{T}e) = \frac{C_e(\phi \cap \mathbf{T}e)}{C_e(\mathbf{T}e)} = \frac{\sum_{f \in \mathcal{E}} C(\phi \cap \mathbf{T}e \mid \mathbf{T}f) \cdot C(\mathbf{T}f \mid \mathbf{U}e)}{\sum_{f \in \mathcal{E}} C(\mathbf{T}e \mid \mathbf{T}f) \cdot C(\mathbf{T}f \mid \mathbf{U}e)}$$

$C(\mathbf{T}e \mid \mathbf{T}f) = 1$ if $e = f$ and 0 otherwise, and $C(\phi \cap \mathbf{T}e \mid \mathbf{T}e) = C(\phi \mid \mathbf{T}e)$, so

$$C_e(\phi \mid \mathbf{T}e) = \frac{C(\phi \mid \mathbf{T}e) \cdot C(\mathbf{T}e \mid \mathbf{U}e)}{C(\mathbf{T}e \mid \mathbf{U}e)} = C(\phi \mid \mathbf{T}e)$$

Therefore,

$$\sum_{e \in \mathcal{E}} C_e(\phi \mid \mathbf{T}e) \cdot C(\mathbf{T}e) = \sum_{e \in \mathcal{E}} C(\phi \mid \mathbf{T}e) \cdot C(\mathbf{T}e) = C(\phi)$$

□

If you are immodest, EXCONDI reduces to HILD and SCHOENFIELD's EVCONDI. For, if you are immodest, $C(\mathbf{U}e \mid \mathbf{T}f)$ will be 1 if $e = f$ and 0 otherwise, so

$$C_e(\phi) = \sum_{f \in \mathcal{E}} C(\phi \mid \mathbf{T}f) \cdot C(\mathbf{T}f \mid \mathbf{U}e)$$

[34] Because we are considering experiential updates and not certainty updates, the principle I derive here differs from the principle NEW RAT-REF introduced earlier. The only difference is that I have exchanged the total certainty evidence operator $\mathbb{T}$ for the total experiential evidence operator $\mathbf{T}$.

	$s \cap \neg d \cap \mathbf{T}s$	$\neg s \cap \neg d \cap \mathbf{T}\neg s$	$s \cap d \cap \mathbf{T}\top$	$\neg s \cap d \cap \mathbf{T}\top$
$\mathbf{U}s$	8/20	0	1/20	1/20
$\mathbf{U}\neg s$	0	8/20	1/20	1/20
	4/10	4/10	1/10	1/10

(a)

	$s \cap \neg d \cap \mathbf{T}s$	$\neg s \cap \neg d \cap \mathbf{T}\neg s$	$s \cap d \cap \mathbf{T}\top$	$\neg s \cap d \cap \mathbf{T}\top$
$\mathbf{U}s$	16/20	0	1/20	1/20
$\mathbf{U}\neg s$	0	0	1/20	1/20
	8/10	0	1/10	1/10

(b)

FIGURE 8: Given the prior credal state shown in figure 8a, the credal state shown in figure 8b is the result both of updating with HCONDI on $\{<\neg d, s>, <d, \top>\}$, and of updating with EXCONDI on s.

$$= C(\phi \mid \mathbf{T}e) \cdot \underbrace{C(\mathbf{T}e \mid \mathbf{U}e)}_{1} + \sum_{f \neq e} C(\phi \mid \mathbf{T}f) \cdot \underbrace{C(\mathbf{T}f \mid \mathbf{U}e)}_{0}$$
$$= C(\phi \mid \mathbf{T}e)$$

As we saw above, EVCONDI entails both INTERNALISM and CONDI. So EXCONDI says: if you are immodest, then your certainty experiment will form a partition, and you should update by conditioning on the true member of this partition.

In other special cases, EXCONDI reduces to HCONDI. Assume your pre-experimental credence that you've been drugged, d, is 20%, your credence that you are drug-free, $\neg d$, is 80%, you think the pear is as likely to taste sweet to you, s as not, $\neg s$, and you think that whether the pear tastes sweet to you is independent of whether or not you're drugged. If you're drug-free and the pear tastes sweet to you, then experience will teach that it does (and no more), $\mathbf{T}s$. If you're drug-free and the pear doesn't taste sweet, then experience will teach that it doesn't (and no more), $\mathbf{T}\neg s$. If you're drugged, then your experience will teach nothing at all, $\mathbf{T}\top$. If you're drug free, then you'll update on whatever experience teaches. However, if you're drugged, you'll incorrectly update on either s or $\neg s$ (with equal probability), even though experience won't teach you either. Then, your pre-experimental credences are as shown in figure 8a. You bite into the pear and it tastes sweet to you. The result of updating on s with EXCONDI is shown in figure 8b. This is exactly the result of updating on the input $\{<\neg d, s>, <d, \top>\}$ with HCONDI.

More generally, suppose that, if the background theory t is true, then experience will teach exactly one of the propositions in the partition $\{e_1, e_2, \ldots, e_N\}$. If, however, the background theory t is false, then experience will teach nothing at all, though you will still erroneously update on one of e_i. Then, your experiential experiment is $\{e_1, e_2, \ldots, e_N, \top\}$, and for each of the e_i, you are certain that

$\mathbf{T}e_i \leftrightarrow t \cap e_i$,[35] and you are certain that $\mathbf{T}\top \leftrightarrow \neg t$. So the result of updating on e_i with EXCONDI will be:

$$\begin{aligned} C_{e_i}(\phi) &= \sum_{f \in \mathcal{E}} C(\phi \mid \mathbf{T}f) \cdot C(\mathbf{T}f \mid \mathbf{U}e_i) \\ &= \sum_{j=1}^{N} C(\phi \mid t \cap e_j) \cdot C(t \cap e_j \mid \mathbf{U}e_i) \;+\; C(\phi \mid \neg t) \cdot C(\neg t \mid \mathbf{U}e_i) \end{aligned}$$

If the background theory t is true, then you will update to e_i iff e_i is true, so $C(t \cap e_i \mid \mathbf{U}e_i) = C(t \mid \mathbf{U}e_i)$ and $C(t \cap e_j \mid \mathbf{U}e_i) = 0$ if $j \neq i$. So the above reduces to

$$C_{e_i} = C(\phi \mid t \cap e_i) \cdot C(t \mid \mathbf{U}e_i) \;+\; C(\phi \mid \neg t) \cdot C(\neg t \mid \mathbf{U}e_i)$$

If whether you update on e_i is independent of whether the background theory is true, $C(t \mid \mathbf{U}e_i) = C(t)$, then EXCONDI will deliver the same result as updating on the input $\{< t, e_i >, < \neg t, \top >\}$ with HCONDI. If $\mathbf{U}e_i$ is not independent of t, then EXCONDI and HCONDI need not agree. This is for the good, since the fact that HCONDI holds fixed your credence in the background theory t was a problem with that update rule. It is not a problem EXCONDI shares. If $C(t \mid \mathbf{U}e_i) > C(t)$, then updating on e_i will confirm the background theory. If $C(t \mid \mathbf{U}e_i) < C(t)$, then updating on e_i will disconfirm the background theory. If the chance of the pear tasting sour is one-in-a-million, while Sabeen is as likely as not to have slipped you the drug, then $C(d \mid \mathbf{U}\neg s)$ will be much greater than $C(d)$, and EXCONDI will set $C_{\neg s}(d)$ much higher than $C(d)$.

Given an experiential experiment $\mathcal{E} = \{e_1, e_2, \dots, e_N\}$ and a pre-experimental credence, EXCONDI tells you which propositions to become certain of upon updating on each e_i. In this way, EXCONDI determines certainty experiments from experiential experiments and pre-experimental credences. Unlike EVCONDI, the certainty experiments EXCONDI determines need not be partitions, so, unlike EVCONDI, EXCONDI is consistent with EXTERNALISM. Even if your experiential experiment is a partition, your certainty experiment need not be. For instance, in our simplified version of Williamson's clock, your experiential experiment, $\{1, 2, 3, 4\}$, was a partition, though, according to EXCONDI, your certainty experiment was the non-partition $\{\neg 1, \neg 2, \neg 3, \neg 4\}$. Similarly, even if your experiential experiment is not a partition, your certainty experiment still could be. Immodesty suffices for a partitional certainty experiment whether your experiential experiment is a partition or not. And both your experiential and certainty experiments could be non-partitions. Consider, for instance, the example from figure 8. Your experiential experiment is the non-partition $\{s, \neg s, \top\}$, while your certainty experiment is

---

[35] I suppose here that the lessons of experience are factive: experience will teach that e_i only if e_i is true.

	$b \cap \mathbf{T}b$	$v \cap \mathbf{T}v$	$g \cap \mathbf{T}g$
$\mathbf{U}b$	7/30	2/30	1/30
$\mathbf{U}v$	2/30	7/30	1/30
$\mathbf{U}g$	1/30	1/30	8/20
	1/3	1/3	1/3

(a)

	$b \cap \mathbf{T}b$	$v \cap \mathbf{T}v$	$g \cap \mathbf{T}g$
$\mathbf{U}b$	49/100	4/100	1/100
$\mathbf{U}v$	14/100	14/100	1/100
$\mathbf{U}g$	7/100	2/100	8/100
	7/10	2/10	1/10

(b)

FIGURE 9

the non-partition $\{\neg d \to s, \neg d \to \neg s\}$.

It could be that, though experience will teach you one of a partition of propositions, this will not allow you to become certain of any new proposition. You observe at a cloth in dim lighting. You know the cloth is either blue, violet, or green, you regard each color as equally likely, and while you know that experience will teach you its true color, you are not certain to correctly learn experience's lesson. If the cloth is blue, you think you're most likely to recognize it as blue (70%), though you may incorrectly think it's violet (20%) or green (10%). Similarly, if it's violet, you'll most likely recognize it as violet (70%), though you may think its blue (20%) or green (10%) instead. And, if it's green, you'll most likely recognize it as green (80%), though perhaps you'll mistake it for blue or violet (10% each). Then, your credences, before looking at the cloth, are as shown in figure 9a. The result of updating this pre-experimental credence distribution on b with EXCONDI is shown in figure 9b. Your credences shift along the partition $\{b, v, g\}$, though you do not become certain of any proposition.

Learning episodes like this were discussed by JEFFREY (1965) (though JEFFREY thinks about them differently from the way I am suggesting we think about them).[36] In JEFFREY's treatment, the input to your update is a set of ordered pairs of propositions e_i and real numbers α_i, $\{< e_i, \alpha_i >\}_i$, such that the e_i form a partition and the α_i sum to 1. Inputs like these are eponymously called 'Jeffrey shifts'. The interpretation of a Jeffrey shift is that α_i is the post-experimental credence that e_i which has been rationalized in experience. In the case of the dimly-lit cloth, your Jeffrey shift may be $\{< b, 7/10 >, < v, 2/10 >, < g, 1/10 >\}$. If so,

[36] See also FIELD (1978), who presents a rule similar to JEFFREY's. FIELD's way of thinking about these learning experiences was also different from JEFFREY's, as well as from the way I am suggesting we think about them here.

then the post-experimental credence distribution in figure 9b is exactly the one recommended by JEFFREY's update rule, known as:

> JEFFREY CONDITIONALIZATION
> If C is your pre-experimental credence function and $\{< e_i, \alpha_i >\}_i$ is the Jeffrey shift acquired in the experiment, then your rational post-experimental credence function, C^+, is such that, for every proposition ϕ,
>
> (JCONDI) $$C^+(\phi) = \sum_i C(\phi \mid e_i) \cdot \alpha_i$$

More generally, suppose that 1) experience will teach one of the partition $\{e_i\}_i$, and 2) these lessons are factive, so that experience teaches e_i only if e_i is true. Then, updating with EXCONDI on the proposition $e \in \{e_i\}_i$ is equivalent to updating with JCONDI on the Jeffrey shift $\{< e_i, C(\mathbf{T}e_i \mid \mathbf{U}e) >\}_i$.

Proof. By definition, $\{\mathbf{T}e_i\}_i$ is a partition. If $\mathbf{T}e_i$ entails e_i, for each e_i, and if $\{e_i\}_i$ is a partition, then $\{\mathbf{T}e_i\}_i$ and $\{e_i\}_i$ must be the very same partition. Then, $C(\phi \mid \mathbf{T}e_i) = C(\phi \mid e_i)$. So EXCONDI says:

$$\begin{aligned} C_e(\phi) &\stackrel{!}{=} \sum_i C(\phi \mid \mathbf{T}e_i) \cdot C(\mathbf{T}e_i \mid \mathbf{U}e) \\ &= \sum_i C(\phi \mid e_i) \cdot C(\mathbf{T}e_i \mid \mathbf{U}e) \end{aligned}$$

which is the result of updating on the Jeffrey shift $\{< e_i, C(\mathbf{T}e_i \mid \mathbf{U}e) >\}_i$. $\square$

To repeat: the way that JEFFREY thought about Jeffrey shifts is very different from the way that I am suggesting we think about them. However, it is still noteworthy that EXCONDI provides us with a way of understanding experiments which rationalize certainty in no proposition; and that, equipped with that understanding, the update prescribed by EXCONDI aligns with the update prescribed by JCONDI, given a Jeffrey shift on the natural partition. Insofar as the prescriptions of JCONDI were plausible in these kinds of experiments, this lends credence to EXCONDI.

The same goes for EVCONDI and HCONDI. HILD and SCHOENFIELD did not explicitly understand their rule as an update strategy for experiential experiments in which you are immodest. Nevertheless, their arguments for EVCONDI implicitly assume immodesty, and the paradigm cases with which HILD and SCHOENFIELD were concerned may be understood in this way. So understood, EXCONDI will agree with EVCONDI. Insofar as the arguments for and prescriptions of EVCONDI were plausible in conditions of immodesty, this lends credence to EXCONDI. Similarly, I did not explicitly understand HCONDI as an update strategy for experiential experiments in which you stand to learn one of a partition of propositions if some background theory t is true, and you stand to learn nothing if t is false. However, the paradigm cases with which I was concerned may be understood in this way. And, so understood, EXCONDI agrees with HCONDI (*modulo* updating your credence in the theory t). Insofar as the prescriptions of HCONDI were plausible for these cases, this lends credence to EXCONDI.

5 In Summation

The externalist says that our certainty experiments need not form a partition. If the externalist is correct, then conditionalization cannot hold in full generality. If your certainty experiment can fail to form a partition, then conditionalization can advise you to plan to become more confident of a proposition, no matter the experiment's outcome, and this is not a rational plan. I've suggested a way for externalists to understand how a certainty experiment could fail to form a partition, as well as an update strategy for these situations. On this understanding, we begin with the notion of an *experiential* experiment—the set of propositions which might be the strongest proposition experience teaches. Your update is a plan for responding to this experiential experiment. If experience may teach that e and no more, then your update says which post-experimental credence to adopt in this contingency. If you are certain to follow this plan correctly—if you are immodest—then your certainty experiment will form a partition. If, however, you modestly foresee possibilities of error, then your certainty experiment need not form a partition. In either case, you should strategize to update with EXCONDI. This is the available update strategy with maximal expected accuracy (given any strictly proper, additive, and extensional measure of accuracy). If you are immodest, then EXCONDI agrees with HILD and SCHOENFIELD's EVCONDI—which is to say: if you are immodest, then your certainty experiment will form a partition, and EXCONDI agrees with CONDI. If you stand to learn that one of a partition of propositions is true if a background theory t is true (and you stand to learn nothing otherwise), then EXCONDI agrees with HCONDI—so long as the background theory t is independent of how you've updated. Moreover, EXCONDI solves HCONDI's problem with updating your credence in the background theory. If your updating on e makes it more/less likely that t is true, then, unlike HCONDI, EXCONDI will raise/lower your credence that t. If you are modest enough, your experience could fail to rationalize certainty in *any* proposition. These look like the motivating cases for JEFFREY's update rule JCONDI. And, in these cases, EXCONDI agrees with JCONDI, given a Jeffrey shift on the natural partition.

I don't say that externalism is correct. Perhaps it is, perhaps not. (Myself, I'm undecided.) What I say is this: if externalism is correct, it is correct because you should sometimes modestly anticipate the possibility of updating contrary to plan. This modesty could be moderate or extreme. It could be that, though you will certainly adopt a post-experimental credence which was planned for *some* contingency, you may end up adopting the wrong post-experimental credence for the contingency in which you in fact find yourself. This is a moderate modesty. If you are extremely modest, you foresee the possibility of adopting a post-experimental credence which you didn't plan to adopt under *any* contingency. If you are extremely modest, I have no advice to offer. If, however, you are moderately modest, then I say: you should plan to update with EXCONDI.

A Technicalities

Proposition 1. EXCONDI *is the available mixed update strategy which maximizes expected accuracy, given any strictly proper, additive, and extensional measure of accuracy.*

Proof. If $\mathscr{A}$ is additive, then $\mathscr{A}(C_e, w)$ has the form $\sum_\phi \mathscr{A}(C_e(\phi), \phi, w)$, and the expected accuracy of an available mixed update strategy μ is given by

$$\begin{aligned}&\sum_{f\in\mathcal{E}}\sum_{w\in\mathbf{T}f} C(w)\cdot\sum_{e\in\mathcal{E}} C(\mathbf{U}e\mid\mathbf{T}f)\cdot\sum_{\phi}\mathscr{A}(C_e(\phi),\phi,w)\\ =&\sum_{\phi}\sum_{e\in\mathcal{E}}\sum_{f\in\mathcal{E}} C(\mathbf{U}e\mid\mathbf{T}f)\cdot\sum_{w\in\mathbf{T}f} C(w)\cdot\mathscr{A}(C_e(\phi),\phi,w)\end{aligned}$$

Pick a ϕ and pick an $e\in\mathcal{E}$. Let $x := C_e(\phi)$. We want the choice of x which maximizes the equation above. This will be the choice which maximizes

$$\sum_{f\in\mathcal{E}} C(\mathbf{U}e\mid\mathbf{T}f)\cdot\sum_{w\in\mathbf{T}f} C(w)\cdot\mathscr{A}(x,\phi,w)$$

If $\mathscr{A}$ is extensional, then there is some $\mathscr{A}_1$ and some $\mathscr{A}_0$ such that

$$\mathscr{A}(x,\phi,w)=\begin{cases}\mathscr{A}_1(x) & \text{if } w\in\phi\\ \mathscr{A}_0(x) & \text{if } w\notin\phi\end{cases}$$

So the choice of x with maximal expected accuracy will be the one which maximizes

$$\begin{aligned}&\sum_{f\in\mathcal{E}} C(\mathbf{U}e\mid\mathbf{T}f)\left(\sum_{w\in\mathbf{T}f\cap\phi} C(w)\cdot\mathscr{A}_1(x) + \sum_{w\in\mathbf{T}f\cap\neg\phi} C(w)\cdot\mathscr{A}_0(x)\right)\\ =&\sum_{f\in\mathcal{E}} C(\mathbf{U}e\mid\mathbf{T}f)(\mathscr{A}_1(x)\cdot C(\mathbf{T}f\cap\phi) + \mathscr{A}_0(x)\cdot C(\mathbf{T}f\cap\neg\phi))\\ =&\mathscr{A}_1(x)\left(\sum_{f\in\mathcal{E}} C(\mathbf{T}f\cap\phi)\cdot C(\mathbf{U}e\mid\mathbf{T}f)\right) + \mathscr{A}_0(x)\left(\sum_{f\in\mathcal{E}} C(\mathbf{T}f\cap\neg\phi)\cdot C(\mathbf{U}e\mid\mathbf{T}f)\right)\end{aligned}$$

If a choice of x maximizes this equation, then it will continue to maximize it if we divide it by the positive constant $C(\mathbf{U}e)$:

$$\begin{aligned}&\mathscr{A}_1(x)\left(\sum_{f\in\mathcal{E}}\frac{C(\mathbf{T}f\cap\phi)\cdot C(\mathbf{U}e\mid\mathbf{T}f)}{C(\mathbf{U}e)}\right) + \mathscr{A}_0(x)\left(\sum_{f\in\mathcal{E}}\frac{C(\mathbf{T}f\cap\neg\phi)\cdot C(\mathbf{U}e\mid\mathbf{T}f)}{C(\mathbf{U}e)}\right)\\ =&\mathscr{A}_1(x)\left(\sum_{f\in\mathcal{E}}\frac{C(\mathbf{T}f\cap\phi)}{C(\mathbf{T}f)}\cdot\frac{C(\mathbf{U}e\cap\mathbf{T}f)}{C(\mathbf{U}e)}\right) + \mathscr{A}_0(x)\left(\sum_{f\in\mathcal{E}}\frac{C(\mathbf{T}f\cap\neg\phi)}{C(\mathbf{T}f)}\cdot\frac{C(\mathbf{U}e\cap\mathbf{T}f)}{C(\mathbf{U}e)}\right)\\ =&\mathscr{A}_1(x)\left(\sum_{f\in\mathcal{E}} C(\phi\mid\mathbf{T}f)\cdot C(\mathbf{T}f\mid\mathbf{U}e)\right) + \mathscr{A}_0(x)\left(\sum_{f\in\mathcal{E}} C(\neg\phi\mid\mathbf{T}f)\cdot C(\mathbf{T}f\mid\mathbf{U}e)\right)\end{aligned}$$

$\sum_{f\in\mathcal{E}} C(-\mid\mathbf{T}f)\cdot C(\mathbf{T}f\mid\mathbf{U}e)$ is a probability function, so the above may be written as

$$\mathscr{A}_1(x)\cdot\alpha + \mathscr{A}_0(x)\cdot(1-\alpha)$$

with $\alpha := \sum_f C(\phi\mid\mathbf{T}f)\cdot C(\mathbf{T}f\mid\mathbf{U}e)$. Since $\mathscr{A}$ is strictly proper, α is the unique value of x which maximizes the equation above. So, for any e and ϕ, the unique choice of $C_e(\phi)$ which maximizes expected accuracy is

$$C_e(\phi) = \sum_{f \in \mathcal{E}} C(\phi \mid \mathbf{T}f) \cdot C(\mathbf{T}f \mid \mathbf{U}e)$$

□

References

Briggs, Rachael. 2009. "Distorted Reflection." *Philosophical Review*, vol. 118 (1): 59–85. [12]

Christensen, David. 2010. "Rational Reflection." *Philosophical Perspectives*, vol. 24: 121–140. [13]

Elga, Adam. 2007. "Reflection and Disagreement." *Noûs*, vol. 41 (3): 478–502. [14], [15]

—. 2013. "The puzzle of the unmarked clock and the new rational reflection principle." *Philosophical Studies*, vol. 164: 127–139. [13], [14], [15], [28]

Field, Hartry. 1978. "A Note on Jeffrey Conditionalization." *Philosophy of Science*, vol. 45 (3): 361–367. [31]

Gallow, J. Dmitri. 2014. "How to Learn from Theory-Dependent Evidence; or Commutativity and Holism: A Solution for Conditionalizers." *The British Journal for the Philosophy of Science*, vol. 65 (3): 493–519. [1], [2], [18]

Gardner, Martin. 1961. *The Second Scientific American Book of Mathematical Puzzles and Diversions*. Simon and Schuster, New York. [11]

Gibbard, Allan. 2008. "Rational Credence and the Value of Truth." In *Oxford Studies in Epistemology*, Tamar Gendler & John Hawthorne, editors, vol. 2, 143–64. Oxford University Press, Oxford. [22]

Greaves, Hilary & David Wallace. 2006. "Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility." *Mind*, vol. 115 (495): 607–632. [5], [22]

Hall, Ned. 1994. "Correcting the Guide to Objective Chance." *Mind*, vol. 103 (412): 505–517. [15]

Hild, Matthias. 1998a. "Auto-Epistemology and Updating." *Philosophical Studies*, vol. 92: 321–361. [1], [2], [14], [19]

—. 1998b. "The Coherence Argument Against Conditionalization." *Synthese*, vol. 115: 229–258. [1], [2], [14], [19], [20], [21], [22], [28], [32], [33]

Jeffrey, Richard. 1965. *The Logic of Decision*. McGraw-Hill, New York. [1], [31], [32], [33] Joyce, James M. 2009. "Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief." In *Degrees of Belief*, F. Huber & C. Schmidt-Petri, editors, 263–97. Springer, Dordrecht. [22]

Lewis, David K. 1996. "Elusive Knowledge." *Australasian Journal of Philosophy*, vol. 74 (4): 549–567. [22]

—. 1999. "Why Conditionalize?" In *Papers in Metaphysics and Epistemology*, vol. 2, chap. 23, 403–407. Cambridge University Press, Cambridge. [22]

Oddie, Graham. 1997. "Conditionalization, Cogency, and Cognitive Value." *British Journal for the Philosophy of Science*, vol. 48: 533–41. [22]

Pettigrew, Richard. 2012. "An Improper Introduction to Epistemic Utility Theory." In *EPSA Philosophy of Science: Amsterdam 2009*, Henk W. de Regt, Stephan Hartmann & Samir Okasha, editors, 287–301. Springer Netherlands, Dordrecht. doi: 10.1007/978-94-007-2404-4_25. URL `http://dx.doi.org/10.1007/978-94-007-2404-4_25`. [22]

Predd, Joel, Robert Seiringer, Elliot H. Lieb, Daniel N. Osherson, H. Vincent Poor & Sanjeev R. Kulkarni. 2009. "Probabilistic Coherence and Proper Scoring Rules." *IEEE Transations on Information Theory*, vol. 55 (10): 4786–4792. [22]

Salow, Bernhard. forthcoming. "The Externalist's Guide to Fishing for Compliments." *Mind*. [1], [5], [16], [17], [18]

Schoenfield, Miriam. forthcoming. "Conditionalization does not (in general) Maximize Expected Accuracy." *Mind*. [1], [2], [19], [20], [21], [22], [24], [28], [32], [33]

Selvin, Steve. 1975. "A Problem in Probability." *The American Statistician*, vol. 29 (1): 67–71. [11]

Stalnaker, Robert C. 2009. "On Hawthorne and Magidor on Assertion, Context, and Epistemic Accessibility." *Mind*, vol. 118 (470): 399–409. [5]

van Fraassen, Bas C. 1984. "Belief and the Will." *The Journal of Philosophy*, vol. 81 (5): 235–256. [11], [12]

—. 1995. "Belief and the Problem of Ulysses and the Sirens." *Philosophical Studies*, vol. 77: 7–37. [11]

Williamson, Timothy. 2000. *Knowledge and its Limits*. Oxford University Press, Oxford. [2], [3]

—. 2014. "Very Improbable Knowing." *Erkenntnis*, vol. 79 (5): 971–999. [3], [5]