Updating for Externalists

J. Dmitri Gallow †

Abstract

The externalist says that your evidence could fail to tell you what evidence you do or not do have. In that case, it could be rational for you to be uncertain about what your evidence is. This is a kind of uncertainty which orthodox Bayesian epistemology has difficulty modeling. For, if externalism is correct, then the orthodox Bayesian learning norms of conditionalization and reflection are inconsistent with each other. I recommend that an externalist Bayesian reject conditionalization. In its stead, I provide a new theory of rational learning for the externalist. I defend this theory by arguing that its advice will be followed by anyone whose learning dispositions maximize expected accuracy. I then explore some of this theory's consequences for the rationality of epistemic *akrasia*, peer disagreement, undercutting defeat, and uncertain evidence.

Orthodox Bayesian epistemology is designed to model rational uncertainty.[1] When you lack relevant evidence, its norms permit uncertainty about various and sundry matters: the weather, the victor, the price of tea in China. But there is a certain kind of uncertainty which orthodox Bayesianism has more difficulty modeling: uncertainty about what evidence you do or do not possess. Influential arguments in support of orthodox Bayesianism take for granted that your evidence will always tell you what your total evidence is, so that there may be no uncertainty about what your evidence says. Let's call this thesis 'internalism',

Final draft; forthcoming in *Noûs*.

† For helpful conversations and feedback on this material, I am indebted to Sara Aronowitz, Adam Bjorndahl, Catrin Campbell-Moore, Nilanjan Das, Kevin Dorst, Daniel Drucker, Julien Dutant, Adam Elga, Jeremy Goodman, Harvey Lederman, Stephen Mackereth, Alexander Meehan, Jim Joyce, Jim Pryor, Daniel Rothschild, Teddy Seidenfeld, Julia Staffel, Robert Steel, Pablo Zendejas Medina, Snow Zhang, and two anonymous reviewers. Thanks also to audiences at Princeton University, University College London, the *Updating and Experience* conference at Ruhr University, Bochum, the *Formal Epistemology* Seminar at Carnegie Mellon University, and the 2019 Pitt-CMU Graduate conference.

[1] As I'll understand the position, orthodox Bayesianism is committed to at least the following theses: *probabilism*, which says that rational credences are (at least finitely additive) probabilities; *conditionalization*, which says that you should be disposed to learn from your evidence by conditioning on it (see §2), and van Fraassen's principle of *reflection*, which says that your current credences should equal your expectation of your future credences (see §2.1). For each of these assumptions, there are epistemologists justly called 'Bayesian' who deny it; but these assumptions compose a familiar, 'off-the-shelf' Bayesian theory of rationality.

and let's call its negation 'externalism'.[2] The externalist thinks that your evidence could fail to tell you what evidence you have or don't have, in which case, it could be rational for you to be uncertain about what your evidence is.

Both internalism and externalism have able defenders. Myself, I'm undecided. I can feel the force of arguments on both sides. So I won't be defending either position here. Instead, I will be asking: what becomes of orthodox Bayesianism if externalism is correct? And on this question, I am decided. The externalist Bayesian should reject the orthodox learning norm (or *updating rule*) of conditionalization. This is a lesson I've learned from Salow (2018), who teaches that, if an externalist Bayesian follows conditionalization, then they will be capable of engaging in acts of deliberate self-delusion—repeatedly 'learning' from experience in such a way as to raise their rational credence in some proposition as high as they like, even when the proposition is false. It is difficult to see this as rational inquiry. So I recommend that the externalist reject conditionalization. In its stead, I will advance a new theory of rational learning for the externalist. I'll defend this theory by arguing that its advice will be followed by anyone whose learning dispositions maximize expected accuracy. And I'll show that those who follow this theory's advice will be incapable of engaging in deliberate self-delusion.

Assuming evidentialism—that is, assuming that the rationality of your doxastic states is determined by the evidence you possess—externalism entails that your evidence could fail to tell you whether your doxastic states are rational or not. For this reason, externalism has played a starring role in recent debates about the rationality of epistemic *akrasia*. Some externalists have held that your evidence could make it likely both that it will rain and that your evidence doesn't make it likely that it will rain. In that case, they have proposed that it is rational for you to be epistemically *akratic*, believing both that it will rain and that it's irrational to believe that it will rain. Relatedly, some externalists have held that the disagreement of an epistemic peer—who has all the evidence that you do, and is equally good at evaluating it as you are—may give you reason to think that your belief in rain was irrationally formed; nevertheless, this need not give you any reason to revise your views about the weather.[3]

The externalist theory of learning I'll develop here yields a distinctive form of externalism, according to which certain kinds of epistemic *akrasia* are always irra-

---

[2] Internalism is presupposed by the Lewis-Teller *diachronic Dutch book argument* for conditionalization (for more, see Gallow 2017), as well as the more recent accuracy (or 'epistemic utility') arguments for conditionalization from Greaves & Wallace and Briggs & Pettigrew (forthcoming) (for more, see Schoenfield 2017a). For justifications of conditionalization which do not presuppose internalism, see (for instance) van Fraassen (1989, ch. 13), Lange (1999), Leitgeb & Pettigrew (2010b), Titelbaum (2013, ch. 7), Gallow (2019), and Zendejas Medina (ms).

[3] For more, see Elga (2013), Weatherson (ms, 2013), Horowitz (2014), Greco (2014), Lasonen-Aarnio (2014, 2015, forthcoming), and §3.

tional; and, if the disagreement of an epistemic peer gives you reason to think that your beliefs were irrationally formed, then it can be rational for you to 'conciliate' with that peer after learning of the disagreement. It additionally gives guidance in cases of undercutting defeat, and learning experiences in which certainty in no proposition has been rationalized. The latter cases are usually treated with *Jeffrey conditionalization* (see JEFFREY, 1965). My past self treated the former with a norm I called *holistic conditionalization* (GALLOW, 2014). I will show that, in the paradigm cases, the theory defended here agrees with both of these learning norms.

1 Internalism and Externalism

In general, to be an internalist is to think that some condition must always lie within your epistemic reach. Given some condition c, an internalist says: if you satisfy c, then you must have access to the fact that you satisfy c. An externalist, in contrast, says that you may satisfy c without having access to the fact that you satisfy c. Different conditions and different kinds of access yield different forms of internalism and externalism. For instance: let the condition be *being in pain*, and say that you have access to a fact when you know it. We then get the internalist thesis that, if you are in pain, you must know that you are in pain, and the corresponding externalist thesis that you may be in pain without knowing that you are in pain.

To get the form of internalism that I'll be interested in, let the condition be *possessing the total evidence e*, and say that you have access to a fact when it is part of your evidence. Then, the internalist says: whenever e is your total evidence, your evidence must say that e is your total evidence. The externalist: on the contrary, sometimes e can be your total evidence without your evidence telling you that e is your total evidence.[4]

> INTERNALISM
> If e is your total evidence, then your evidence must tell you that e is your total evidence.
>
> $$\Box(\mathbf{T}e \rightarrow \mathbf{ET}e)$$

> EXTERNALISM
> You may have the total evidence e without your evidence telling you that e is your total evidence.
>
> $$\Diamond(\mathbf{T}e \wedge \neg\mathbf{ET}e)$$

[4] Throughout, 'e', 'f', and 'ϕ' are meta-variables ranging over propositions. I'll use 'e' and 'f' when I'm presupposing that the proposition is potentially your evidence. I'll use 'ϕ' when I'm not making this presupposition.

Throughout, I'll use $\ulcorner \mathbf{E}e \urcorner$ to mean that your evidence says (at least) that e, and $\ulcorner \mathbf{T}e \urcorner$ to mean that your evidence tells you e and no more (that is: that e is your *total* evidence).

We may provide a semantics for the operators **E** and **T** with Kripke models. On that semantics, **E** is a familiar necessity modal—$\ulcorner \mathbf{E}e \urcorner$ is true at a possible world w iff $\ulcorner e \urcorner$ is true at all worlds to which w bears an accessibility relation, R. **T** is less familiar, but its semantics is simple enough: $\ulcorner \mathbf{T}e \urcorner$ is true at w iff $\ulcorner e \urcorner$ is true at *all and only* worlds to which w bears R. Let's assume that evidence must be *consistent*, so that, if your evidence says that e, then your evidence must not also say that $\neg e$: $\Box(\mathbf{E}e \rightarrow \neg\mathbf{E}\neg e)$. Then, internalism is equivalent to the conjunction of the *Positive Access* principle, $\Box(\mathbf{E}e \rightarrow \mathbf{EE}e)$, and the *Negative Access* principle, $\Box(\neg\mathbf{E}e \rightarrow \mathbf{E}\neg\mathbf{E}e)$. *Positive Access* says that your evidence always tells you what evidence you have. *Negative Access* says that your evidence always tells you what evidence you *don't* have.[5]

Because internalism entails both *Positive Access* and *Negative Access*, an argument against either is an argument against internalism. Since externalism is just the negation of internalism, an argument against either *Positive* or *Negative Access* is an argument for externalism.

Williamson argues for externalism in just this way. He contends that cases of perceptual illusion counterexample *Negative Access*. 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, your evidence doesn't tell you that the wall is red, nor does it rule out that you are in the good case. In the good case, your evidence does tell you that the wall is red. So, in the bad case, your evidence does not tell you that you don't have the evidence that the wall is red. Even so, you don't have this evidence. So, in the bad case, $\neg\mathbf{E}r \wedge \neg\mathbf{E}\neg\mathbf{E}r$, where '$r$' says that the wall is red. So *Negative Access* is false. The internalist could deny that, in the good case, your evidence tells you that the wall is red. Rather, they may say, in both the good and the bad case, your evidence merely tells you that the wall appears red, and/or that you believe the wall is red. Alternatively, they could say: in the bad case, even though the wall isn't red, your evidence still tells you that the wall is red.

Williamson additionally argues that cases in which your perceptual knowledge is inexact counterexample *Positive Access*. 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 lo-

[5] If we assume that your evidence is *factive*—that, if your evidence tells you that e, then e must be true, $\Box(\mathbf{E}e \rightarrow e)$—then *Negative Access* entails *Positive Access*. But if we merely assume that evidence is *consistent*, *Negative* and *Positive Access* are logically independent.

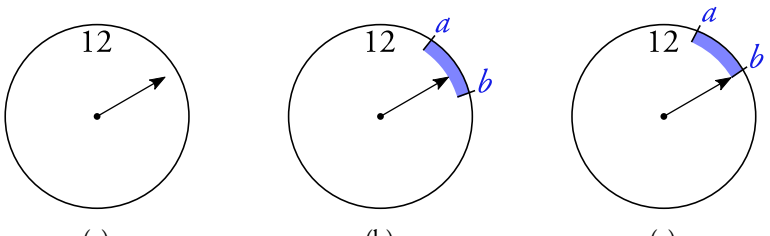


FIGURE 1: A distant and brief glimpse at the unmarked clock (1a) provides the evidence that the clock hand is positioned within some interval of values $[a, b]$ (1b); and you've learned that, if the clock hand is positioned at b, then the glimpse does not provide the evidence that the clock hand is no further than b (1c). These assumptions contradict the *Positive Access* principle, $\Box(\mathbf{E}e \rightarrow \mathbf{EE}e)$.

cation of the clock hand.[6] At most, you learn that the clock hand is located in some interval (see figure 1b). Grant also that your evidence will leave a 'margin-for-error', so 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 *Positive Access*. 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 most your evidence tells you about the position of the clock hand is that it lies in some interval $[a, b]$, with $a < b$.

A2) Your evidence says that: if the clock hand is located at b, then your evidence won't tell you that it is located no further than b (since your evidence must leave a margin-for-error).

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

A3) Your evidence tells you what evidence you have.

$$\mathbf{E}e \rightarrow \mathbf{EE}e$$

To see that these three claims are inconsistent, note that we can get (A4) by contraposition on (A2):

(A4) $$\mathbf{E}[\mathbf{E}(H \leqslant b) \rightarrow H \neq b]$$

Assuming that the evidence operator $\mathbf{E}$ satisfies the K-axiom ($\mathbf{E}(\phi \rightarrow \psi) \rightarrow$

[6] Throughout, I will use ⌜learn that e⌝ to mean ⌜acquire evidence which tells you that e⌝.

$(\mathbf{E}\phi \rightarrow \mathbf{E}\psi)$), (A4) entails (A5).

(A5) $$\mathbf{EE}(H \leqslant b) \rightarrow \mathbf{E}(H \neq b)$$

(A1) entails (A6).

(A6) $$\mathbf{E}(H \leqslant b)$$

From (A6) and (A3), we have

(A7) $$\mathbf{EE}(H \leqslant b)$$

And from (A7) and (A5),

(A8) $$\mathbf{E}(H \neq b)$$

But (A8) 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.[7]

So (A1), (A2), and (A3) are inconsistent. WILLIAMSON thinks that (A3) is the least plausible of the three; but others, like SALOW (2018) and STALNAKER (2009), choose instead to reject (A2) and retain *Positive Access*.[8]

WILLIAMSON's unmarked clock is a nice example of the kinds of cases externalists take to be possible, if not commonplace. However, the example is more complicated than it needs to be for my purposes. So let me introduce a simplified model of the unmarked clock. In this model, the clock hand may point at one of four positions: 1, 2, 3, or 4 (See figure 2). If it points at 1, then, since your evidence must leave a margin-for-error, it will leave open that it points at 4 or 2 instead, and your *total* evidence will just be that it does *not* point at 3 (figure 2a). Likewise, if it points at 2, then, since your evidence must leave a margin-for-error, your total evidence will be that it does *not* point at 4 (figure 2b). In general, your total evidence will be that the clock hand is not at the position opposite its actual position. This model is simplistic and psychologically implausible, but the lessons we learn from it will carry over to the more realistic cases, so I'll continue to focus on this simplified model throughout.

Let me introduce some (stipulative) terminology. I'll call a set of proposi-

[7] 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: $\mathbf{E}[H = b - \epsilon \rightarrow \neg\mathbf{E}(H < b)]$. (The reasoning is exactly the same as in the body, *mutatis mutandis*.)

[8] STALNAKER (2009) does not explicitly discuss the positive access principle for *evidence*; his focus is the positive access for *rational belief*: if it's rational to believe that ϕ, then it's rational to believe that it's rational to believe that ϕ.

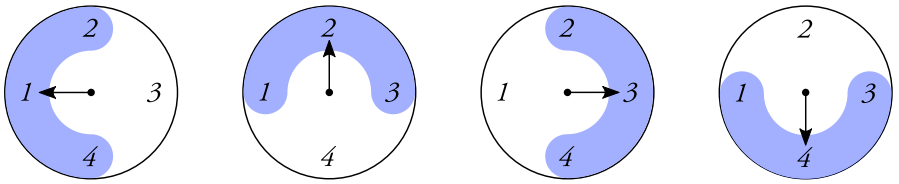


FIGURE 2: A simplified model of Williamson's unmarked clock. The clock hand could point at position 1, 2, 3, or 4. If it points at 1, your total evidence will be that it's not at 3 (2a). If it points at 2, your total evidence will be that it's not at 4 (2b), and similarly for positions 3 and 4. Your experiment, then, is $\mathcal{E} = \{\neg 3, \neg 4, \neg 1, \neg 2\}$.

tions, $\mathcal{E} = \{e_1, e_2, \ldots, e_N\}$, an *experiment*. Intuitively, the set $\mathcal{E}$ contains all and only the propositions which may be your total evidence. Let's say (again, as a terminological stipulation) that you are *conducting the experiment* $\mathcal{E}$ iff: for each $e_i \in \mathcal{E}$, e_i may be your total evidence, and, moreover, your total evidence *must* be one of the $e_i \in \mathcal{E}$. 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. 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.[9] In our simplified model of WILLIAMSON's clock, you will either acquire the total evidence $\neg 3$ (you'll learn this iff 1 is true), the total evidence $\neg 4$ (which you'll learn iff 2 is true), the total evidence $\neg 1$ (iff 3 is true), or the total evidence $\neg 2$ (iff 4 is true).[10] So your experiment is the set $\{\neg 4, \neg 3, \neg 2, \neg 1\}$.

2 UPDATING

Let me assume that you have opinions about how likely various propositions are. I'll call these opinions of yours *credences*, and I'll represent them with a function, C (for 'credence'), from propositions to numbers between 0% and 100%. The interpretation is that $C(\phi)$ represents how likely you take the proposition ϕ to be. I'll assume throughout that, if you are rational, then C will be a probability function.[11] I will also assume that, in addition to these credences, you have

[9] I borrow this terminology from GREAVES & WALLACE (2006).

[10] Here, I am using 'n' to stand for the proposition that the clock hand points at position n—a convention I'll continue to follow throughout.

[11] In general, your opinions may not be defined over a (σ-)*algebra*—a set of propositions containing the tautology, $\top$, and closed under negation and (countable) union. The usual probability axioms assume that C is defined over a set of propositions like this. But we may be less demanding and count your opinions as probabilistic so long as *there is* some probability, defined over a full (σ-)algebra, which agrees with you wherever you are opinionated.

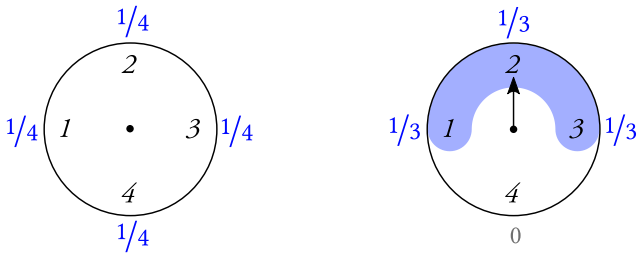


FIGURE 3: In figure 3a, your current credences. You think the hand is equally likely to be at any of the four positions. In figure 3b, the credences CONDI says you should be disposed to adopt upon learning that ¬4 (and no more). You should become certain that ¬4, and you should think that the hand is equally likely to be at any of the remaining positions.

learning dispositions to revise or update your credences in the light of the evidence $e \in \mathcal{E}$. Let's model these learning dispositions with a function, D (for 'disposition'), from propositions which might be your total evidence, $e \in \mathcal{E}$, to new credence functions. The interpretation is that D_e (the output of the function D, given the input e) is the credence function you are disposed to adopt, if your total evidence is e.[12] I will think about these dispositions the same way I think about dispositions more generally: an individual's disposition may be characterized by a certain *stimulus* condition, and a certain *response* which the individual is disposed to manifest in the stimulus condition. For you to have the learning dispositions represented by D is for you to be disposed to adopt the new credence function D_e in the stimulus condition of having your evidence tell you e (and no more). Therefore, for each $e \in \mathcal{E}$, I'll assume that, in all possibilities in which your total evidence is e, you will adopt the credence function D_e.

Which learning dispositions are rational? How should you be disposed to learn from your evidence? The orthodox Bayesian answer to this question is: you should be disposed to learn from the total evidence e by *conditioning* on e. For instance, in our simplified model of Williamson's clock, suppose that you start out thinking the hand is equally likely to be at any position—as in figure 3a. Upon receiving the total evidence ¬4, conditionalization says that you should become certain that ¬4, and you should think that each of the remaining positions are equally likely—as in figure 3b. Likewise, if you learn that ¬1, ¬2, or ¬3 instead, you should be disposed to become certain in your total evidence, and continue to think that the remaining positions are equally likely. In general, conditionalization says:

[12] Elsewhere, functions like these are referred to as 'epistemic acts', 'strategies', 'plans', and 'credal gambles'.

> CONDITIONALIZATION
> Be disposed to respond to the total evidence e by adopting your current credence function, C, *conditioned on* e.

(CONDI) $$D_e(\phi) \stackrel{!}{=} C(\phi \mid e)$$

(I place an exclamation over an equals sign to say that the equality ought to hold. CONDI does not claim that $D_e(\phi)$ *will* be $C(\phi \mid e)$—it says instead that it *should* be.)

I've come to believe that the externalist should not endorse CONDI in full generality, for at least two reasons: firstly, because externalist conditionalizers must accept the rationality of *deliberate self-delusion* (§2.1); and, secondly, because if externalism is correct, then pursuing accurate credences will lead you to violate CONDI (§2.2).

2.1 Externalism, Conditionalization, and Self-Delusion

Return to our simplified model of Williamson's clock, and suppose that, before looking, a reliable source informs you that the clock hand is not at position 4. Learning this also teaches you that a glimpse at the clock won't teach you that $\neg 2$. However, since the clock hand could still point at either 1, 2, or 3, a glimpse at the clock could still teach either $\neg 3, \neg 4$, or $\neg 1$. (Of course, you already know $\neg 4$, so if that's what your glimpse tells you, you won't learn anything new.) So, in taking a glimpse at the clock, you will be conducting the experiment $\mathcal{E} = \{\neg 3, \neg 4, \neg 1\}$.

If, before looking, you think the clock hand is just as likely to be at 1 as it is to be at 2 or 3, then your credences will be as shown in figure 4a. If you learn $\neg 3$ (and no more) and you condition on this evidence, then your credence that 2 will rise to $1/2$, as shown in figure 4b. Likewise, if you learn $\neg 1$ (and no more) and you condition on this evidence, then your credence that 2 will rise to $1/2$, as shown in figure 4d. If, on the other hand, you learn $\neg 4$ (no more) and you condition on this, then since you were already certain of $\neg 4$, your credence in 2 will remain unchanged; it will stay put at $1/3$, as shown in figure 4c.

So, when conducting the experiment $\{\neg 3, \neg 4, \neg 1\}$, if you are disposed to condition on your total evidence, then your credence that 2 is guaranteed to not fall, and it may rise. Notice also that, in advance, you can recognize that your credence that 2 will rise if and only if the clock hand is *not* at 2. Of course, rational learning dispositions may end up taking you further from the truth in some circumstances. What's going on here is more disturbing than that. It's not just that these learning dispositions *may* take you further from the truth—they are actually quite *likely* to do so. Moreover, your updated credence that 2 is *anti-correlated* with the truth about whether 2. And you are certain in advance to end up *no closer* to the truth about whether 2.

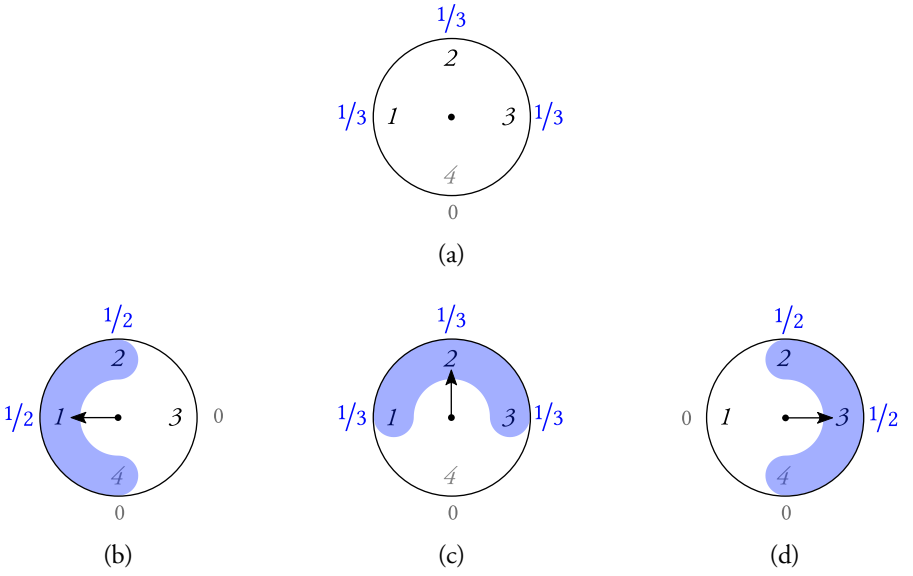


FIGURE 4: In figure 4a, your current credences. In figure 4b, the credences CONDI says you should be disposed to adopt upon learning that ¬3 (and no more). In figure 4c, the credences CONDI says you should be disposed to adopt upon learning that ¬4 (and no more). And, in figure 4d, the credences CONDI says you should be disposed to adopt upon learning that ¬1 (and no more).

SALOW (2018) draws our attention to another disturbing feature of these learning dispositions. With learning dispositions like these, you could engage in a kind of *intentionally biased inquiry*—inquiry which you expect to raise your rational credence in some proposition, even when that proposition is false. 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 credence that p is $1/3$ (though the precise value won't matter). Then, if your learning dispositions conform to CONDI, here's a recipe for raising your credence that p: first, tell a confidant who knows the truth about p to place the clock hand at position 2 iff p is true. If p is false, then they should flip a coin to decide between position 1 and 3. Then, you will be conducting the experiment $\mathcal{E} = \{\neg 3, \neg 4, \neg 1\}$, and, before looking, you will have the credences from figure 4a. You'll think you're $2/3$rds likely to end up 50% sure that the clock hand is at position 2. Since you're certain that $2 \leftrightarrow p$, you'll think that you're $2/3$rds likely to end up thinking p is 50% likely. In fact, you are certain in advance that, so long as p is *false*, you'll end up 50% sure that p is true. And there's no reason this experiment need be conducted only once. Run through the whole exercise again, and you could raise yours credence that p to $2/3$, and—why not?—again, raising it to $4/5$, and again, raising it to $8/9$, and so on and so forth. So long as p is false, you can end up with a credence in p which is as high as you like. (And there's no danger of *lowering* your credence in p. If p is true, you'll still end up with a credence of $1/3$ in p.)

It is incredibly difficult to see this as rational inquiry. Designing this experiment and adopting these learning dispositions is an act of deliberate self-delusion, not a rational search for truth. Let us lay this down as a principle.[13]

> NO SELF-DELUSION
> If you are disposed to raise your credence that ϕ in response to some potential evidence which you'll learn with positive probability, then you must also be disposed to lower your credence that ϕ in response to some potential evidence.

From my perspective, NO SELF-DELUSION is non-negotiable.[14] You may be inclined to disagree because you think that rational agents can be deluded by their past selves. For instance: Cypher decides to enter the matrix and erase his memories of the world outside. He intentionally deludes himself about the external world. But, even so, the beliefs he adopts once inside the matrix, with his memories erased, are the rational ones to adopt given his evidence. I agree, but I don't think this puts any pressure on the principle NO SELF-DELUSION. We should distinguish the *deluded* Cypher inside the matrix from the *deluding* Cypher who makes the decision to put his future self there. We may forgive the deluded Cypher without forgiving his deluding past self. Similarly, we should distinguish your *pre-experimental* self, who has designed this experiment and is disposed to become more confident that p iff p is false, and your *post-experimental* self, who has updated on their evidence and is now more confident that p is true. By analogy with Cypher, you may suggest that your post-experimental self is rational, even though your pre-experimental self is not. I would disagree—Cypher has forgotten how he got there, you have not, and this difference makes a difference with respect to epistemic rationality. However, even if your suggestion is granted, it does not conflict with NO SELF-DELUSION. This principle is only evaluating the learning dispositions of your *pre-experimental* self; it says nothing about the opinions of your *post-experimental* self. What it condemns are dispositions to learn from your evidence which give some proposition a chance of being confirmed while simultaneously protecting it from ever being *dis*confirmed.[15]

---

[13] A very similar principle is called 'Disconfirmability' in WHITE (2006, p. 544), where it is attributed to an early draft of PRYOR (2004). If we assume internalism, and that you're certain to update on your total evidence, then this principle follows from CONDI.

[14] Why require that the evidence which raises your credence that ϕ have positive probability? Suppose a number will be randomly selected from the unit interval and you'll learn its true value. Let ϕ be the proposition that the number will be $1/\sqrt{2}$. Then, your current credence that ϕ will be zero, and it will not get any lower no matter what you learn, but if you learn that ϕ is true, your credence that ϕ will jump up to 100%. This would be a counterexample to NO SELF-DELUSION if we didn't require that you have a positive probability of learning the evidence mentioned in the antecedent.

[15] Thanks to an anonymous reviewer for prompting me to clarify this point.

So long as EXTERNALISM commits us to the possibility of designing experiments like $\mathcal{E} = \{\neg 3, \neg 4, \neg 1\}$, EXTERNALISM, CONDITIONALIZATION, and NO SELF-DELUSION are inconsistent.[16] I, for one, am not prepared to renounce NO SELF-DELUSION. Nor is SALOW, who suggests rejecting EXTERNALISM. Perhaps that is the correct lesson to draw. However, I believe that a plausible externalist position is left standing. This is a version of externalism which accepts NO SELF-DELUSION by denying CONDITIONALIZATION. In §2.2, 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 kind of biased inquiry—inquiry which is *guaranteed* to leave you no less confident in some proposition, and leaves a positive probability of you becoming more confident. The reasons we have to call this kind of biased inquiry irrational carry over to inquiries which you merely *expect* to leave you more confident in some proposition. Say that your learning dispositions are biased in favor of a proposition ϕ iff, when you have those learning dispositions, you expect your updated credence that ϕ to be greater than your current credence that ϕ. Similarly, say that your learning dispositions are biased *against* ϕ iff you expect your updated credence that ϕ to be less than your current credence that ϕ. Then, your learning dispositions are *unbiased* iff, for all propositions ϕ, your current credence that ϕ is equal to your expectation of your updated credence that ϕ. Let's use $\ulcorner \mathbf{U}e \urcorner$ to stand for the proposition that you've *updated* on the proposition e—that is: $\ulcorner \mathbf{U}e \urcorner$ says that you have taken your total evidence to be e, and, in response, adopted the new credence function you are disposed to adopt in that stimulus condition.[17] Thus: when your learning dispositions are given by D, $\ulcorner \mathbf{U}e \urcorner$ says that you've taken your evidence to be e and adopted the new credence function D_e in response. To say that your learning dispositions should be unbiased is just to say that, for all propositions ϕ, your expectation of your updated credence in ϕ should equal your current credence in ϕ:

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

---

[16] *Cf.* HILD (1998a,b), who argues that externalism and conditionalization are inconsistent with REFLECTION (see below). Note that NO SELF-DELUSION is a weakening of REFLECTION.

[17] Thus: even if you are disposed to adopt the same credence function when $\mathbf{T}e$ as you are when $\mathbf{T}f$—even if $D_e = D_f$—the propositions that you've updated on e is a different proposition from the proposition that you've updated on f. If you've updated on e, then you must have (in some sense) taken e to be your evidence; whereas, if you've updated on f, then you must have (in some sense) taken your evidence to be f. As I'm understanding it, 'taking your evidence to be e' need not involve any belief that your evidence is e. For instance, you could hold that having an appropriate sub-personal mechanism categorize your experience as representing that e is one way of taking your evidence to be e.

This is VAN FRAASSEN (1984, 1995)'s principle of REFLECTION.[18] (Notice that, as I'm understanding it, REFLECTION is a constraint on your *learning dispositions*.) SALOW's insight is that your learning dispositions will be biased if and only if they violate REFLECTION. Since biased learning dispositions are irrational, REFLECTION is rationally required.

If internalism is correct, and you are certain to update on your total evidence, then REFLECTION follows from CONDI.[19] So the internalist conditionalizer who is certain that they will correctly update on their evidence will always satisfy REFLECTION.[20] Notice that REFLECTION entails NO SELF-DELUSION—if your credence that ϕ has some positive probability of rising and no probability of falling, then your expectation of your new credence that ϕ will exceed your current credence that ϕ. So an internalist conditionalizer who is certain to update on their total evidence will not be capable of engaging in this kind of deliberate self-delusion.

So long as the externalist thinks that experiments like $\{\neg 3, \neg 4, \neg 1\}$ are possible, they must choose between CONDI and REFLECTION. I believe that they should choose REFLECTION. Looking ahead: the externalist update I will propose in §2.2 below will always satisfy the principle of REFLECTION.

2.2 Externalism and the Pursuit of Accuracy

Your credence function encodes your opinions about how likely various propositions are. If we know whether those propositions are true or false, we can ask: how close to the truth did you get? That is, we can ask: how *accurate* were your credences? I'll assume that we have some way of measuring the accuracy of a credence function, given all the facts. Since all the facts are settled by which possible world is actual, I'll assume that we have some measure of the accuracy of the credence function C at the world w: $\mathscr{A}(C, w)$. There are several accuracy measures which have been defended in the recent literature.[21] However, for my purposes here, the only thing I need assume about the measure $\mathscr{A}$ is that it has

---

[18] VAN FRAASSEN also makes the stronger claim that $C(\phi \mid \mathbf{U}e) \stackrel{!}{=} D_e(\phi)$. The principle I'm calling 'REFLECTION' follows from this claim and the law of total probability. As I'll understand REFLECTION, it governs your dispositions to learn from the evidence acquired in experiments which you are about to conduct, and during which you are certain to not lose evidence. So understood, REFLECTION escapes many of the usual counterexamples (see BRIGGS (2009) for a nice taxonomy of the counterexamples). Those which remain involve credences *de se et nunc*. Credence *de se et nunc* will require us to reject or qualify REFLECTION. Still, I'll ignore these complications for the nonce.

[19] If internalism is true, then you are certain that $e \leftrightarrow \mathbf{T}e$, so $C(\phi \mid e) = C(\phi \mid \mathbf{T}e)$. And if you are certain to update on your total evidence, then $C(\mathbf{U}e) = C(\mathbf{T}e)$. So, if your learning dispositions conform to CONDI, then $\sum_e D_e(\phi) \cdot C(\mathbf{U}e) = \sum_e C(e \mid \mathbf{T}e) \cdot C(\mathbf{T}e) = C(\phi)$.

[20] *Cf.* WEISBERG (2007). In §2.2 below, I'll suggest that there's a close connection between externalism and a modest uncertainty about whether or not you will update on your total evidence.

[21] For more, see JOYCE (1998, 2009) and PETTIGREW (2016a).

the properties explained in this note.[22] If $\mathscr{A}$ has these properties, then I'll say that it is a 'nice' measure of accuracy. All of the accuracy measures which have been defended in the recent literature will count as nice, in this sense.

Once we have a measure of accuracy, we can use it to ask about the *expected* accuracy of your learning dispositions. That is: we may ask how accurate you expect your credences to be, once they've been updated. I'll make the normative assumption that we can evaluate learning dispositions in terms of their expected accuracy, and, in particular, that learning dispositions are rational if they *maximize* expected accuracy. That is: I'll suppose that learning dispositions are rational if, from your current perspective, they are the ones which it would make the most sense for you to adopt, were you concerned only with the accuracy of your credences.[23] In making this assumption, I am allying myself with so-called 'accuracy first' epistemologists, who wish to derive all seemingly evidential epistemic norms from the imperative to rationally pursue accuracy.[24] This is a controversial allegiance. There are many worries we could raise about accuracy-first epistemology.[25] Still, I think the accuracy-first project is an ambitious and compelling research program, and I think it is well worth exploring the kind of externalism which it produces.

Greaves & Wallace (2006) showed that, if we assume that internalism is correct, evidence is factive, and accuracy is measured nicely, then the learning dispositions which maximize expected accuracy are just the ones prescribed by condi. But what if externalism is correct? In this case, Schoenfield (2017a) showed that, so long as accuracy is measuredly nicely, you will maximize expected accuracy iff you are disposed to condition, not on your total evidence, but rather on the proposition *that it is* your total evidence. Let's call this norm *Schoenfield conditionalization*, or just 'scondi'.[26]

---

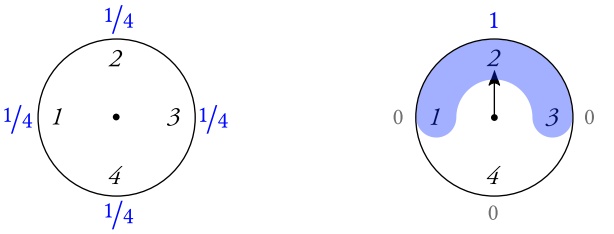
[22] I assume that $\mathscr{A}$ is *additive*, *extensional*, and *strictly proper*. $\mathscr{A}$ is *additive* iff it is of the form $\mathscr{A}(C, w) = \sum_\phi \lambda_\phi \cdot \mathscr{A}(C(\phi), \phi, w)$, for some weights $\lambda_\phi > 0$ of the importance of having an accurate credence in the proposition ϕ and some function $\mathscr{A}(x, \phi, w)$ of the accuracy of a credence x in the proposition ϕ in world w. It is *extensional* iff there are functions $\mathscr{A}_1$ and $\mathscr{A}_0$ such that $\mathscr{A}(x, \phi, w) = \mathscr{A}_1(x)$ if $w \in \phi$ and $\mathscr{A}(x, \phi, w) = \mathscr{A}_0(x)$ if $w \notin \phi$. $\mathscr{A}$ is *strictly proper* iff, for every probabilistic credence function P, the unique credence function C which maximizes $\sum_w P(w) \cdot \mathscr{A}(C, w)$ is P itself.

[23] Though, to be clear, I don't think that, in order for you to count as rational, you have to do anything like *choose* your learning dispositions, or *recognize* that they maximize expected accuracy.

[24] See, *e.g.*, Joyce (1998, 2009), Leitgeb & Pettigrew (2010a,b), Pettigrew (2011, 2012, 2016a,b, 2018), Levinstein (2012), Caie (2013), Easwaran (2013), Bronfman (2014), Schoenfield (2015, 2017b, 2018), and Fitelson et al. (ms).

[25] See, for instance, Berker (2013), Greaves (2013), Carr (2017), Caie (2018), Blackwell & Drucker (2019), and Oddie (2019).

[26] Schoenfield calls this rule 'conditionalization*'. Hild (1998a,b) proposes the same update rule and calls it 'auto-epistemic conditionalization'. Schoenfield (2017a) does not directly discuss *learning dispositions*, but rather *update procedures*.



(a) (b)

FIGURE 5: In figure 5a, your current credences. You think the hand is equally likely to be at any of the four positions. In figure 5b, the credences SCONDI says you should be disposed to adopt upon learning that ¬4 (and no more). You should become certain that **T**¬4—that is, you should become certain that the clock hand is pointed at position 2.

SCHOENFIELD CONDITIONALIZATION
Be disposed to respond to the total evidence e by adopting your current credence function, C, conditioned on $\mathbf{T}e$.

(SCONDI) $$D_e(p) \stackrel{!}{=} C(p \mid \mathbf{T}e)$$

To understand what SCONDI says, return to our simplified model of Williamson's clock. Suppose that you think the clock hand is equally likely to point at any of the four positions (figure 5a). Then, SCONDI says that, upon learning that the hand isn't at position 4 (and no more), you should be disposed to become certain *that you have learned this*, **T**¬4. Since your total evidence will be ¬4 iff the clock hand is at position 2, SCONDI says to become certain that the clock hand is at position 2 (figure 5b).

An externalist should be uncomfortable with this recommendation. Remember, the externalist thinks that, in cases like Williamson's clock, your evidence must leave a *margin-for-error*. It is for this reason that they insist that, if the clock hand is at 2, your evidence must leave it open that it is at position 1 or 3 instead. While SCONDI grants the externalist a margin-for-error when it comes to *evidence*, it denies that there is any margin-for-error when it comes to *rational certainty*.

Because SCONDI says that it is rational for you to condition on $\mathbf{T}e$, it says that it is rational for you to be *certain* that $\mathbf{T}e$. Let's say that some fact is within your epistemic reach if it's rational for you to be certain of that fact. Then, according to SCONDI, what your evidence says is always within your epistemic reach. It is always rational for you to be certain about what your evidence says or doesn't say. This is something an externalist should be uncomfortable saying. Externalists should want to endorse the thesis I'll call *certainty externalism*.

CERTAINTY EXTERNALISM
Your total evidence may be e without it being rational for you to be certain that your total evidence is e

The Williamsonian arguments for externalism carry over straightforwardly to certainty externalism. If anything, those arguments are stronger when transposed into the key of rational certainty. If they establish externalism, then they should likewise establish certainty externalism. So externalists should be certainty externalists. But this means rejecting SCONDI, since SCONDI says that it is always certain what your evidence says.

We've seen that SCONDI follows from externalism, together with the imperative to maximize expected accuracy. For this reason, the foregoing could be viewed as an argument against externalism. If the externalist adopts the dispositions to learn from their evidence which maximize expected accuracy, then they will never be uncertain about what their evidence says. But their externalism should commit them to the possibility of rational uncertainty like this. What is the externalist to do? They could, of course, reject the assumptions of accuracy-first epistemology. For instance, they could say that the *telos* of belief is knowledge, not accuracy, and therefore insist that it can be irrational to do what you expect to get you closest to truth. Alternatively, they could try to motivate accepting externalism about *evidence*, but not externalism about *rational certainty*. Both of these options are available, but I have another suggestion—a suggestion which allows the externalist to maintain a close connection between rational learning dispositions and the rational pursuit of accuracy, without forcing them to give up certainty externalism.

My suggestion is that the externalist reject one of the assumptions made back at the beginning of this section. There, I assumed that in *every* possibility in which the stimulus condition $\mathbf{T}e$ is true, you will manifest the response of adopting the new credences D_e. This implicitly assumes that you take your dispositions to respond to evidence to be *flawless*. It assumes that you foresee no possibility in which your learning dispositions *misfire*—no possibility in which your total evidence is e, but you mistake it for the nearby evidence $f \neq e$, and therefore incorrectly update to the new credence function D_f. If that's so, then you will be certain, in advance, that you'll update on the evidence e iff e is your total evidence, for every $e \in \mathcal{E}$. Some terminology: if $C(\mathbf{U}e \leftrightarrow \mathbf{T}e) = 1$, for every $e \in \mathcal{E}$, then let's say that you are *immodest*. And if you are less than certain to respond correctly to your total evidence—if $C(\mathbf{U}e \leftrightarrow \mathbf{T}e) < 1$, for some $e \in \mathcal{E}$—then let's say that you are *modest*.[27] The learning dispositions recommmended by SCONDI need not maximize expected accuracy if you are modest. So my suggestion to the externalist is this: plead modesty—maintain that rationality permits thinking that you may

[27] Beware: this terminology is slightly idiosyncratic. Others will call you modest if you are less than certain that your *current* credences are rational—let's call this *synchronic* modesty. According to the theory of learning I'll defend below, modesty in my sense will lead to synchronic modesty after you've rationally updated on your evidence.

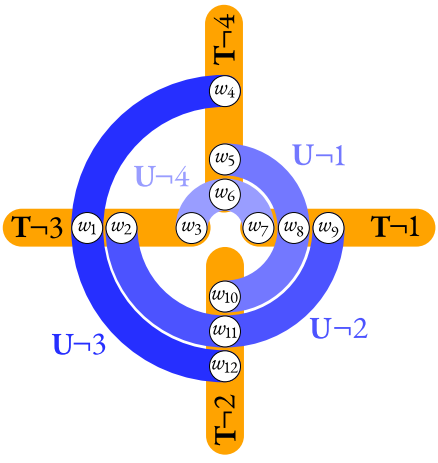


FIGURE 6: w_1, w_2, and w_3 are each possibilities in which your evidence tells you that the clock hand does not point at 3, **T**¬3. At w_1, you correctly update on this evidence, **U**¬3. While, at w_2, you incorrectly update as if your evidence told you that it wasn't at position 2, **U**¬2. And, at w_3, you incorrectly update as if your evidence told you that it isn't at position 4, **U**¬4.

mistake your evidence.[28]

If you think your learning dispositions may misfire, then we should explicitly represent this possibility in our modelling. Return to our simple model of Williamson's clock. Let's complicate things slightly by supposing that, when the clock hand is actually pointing at 1, and therefore, your evidence tells you that it's not pointing at 3, you *may* correctly update on ¬3, but you may also mistakenly respond as if your evidence had told you that it's not pointing at 2, or as if your evidence had told you that it's not pointing at 4. And let's suppose, symmetrically, that when the clock hand is actually pointing at 2, so that your evidence tells you it's not pointing at 4, you *may* correctly update on ¬4, but you may also err by taking your evidence to say either ¬3 or ¬1 instead. And similarly if your evidence tells you ¬1 or ¬2. (See figure 6.)

I'll assume not only that you modestly foresee the possibility of mistaking your evidence. I'll also assume that you have opinions about how likely you are to do so. For instance, in our simple version of Williamson's clock, let's suppose that you think you're 80% likely to correctly update on your evidence, but you think you're 20% likely to err, either clockwise or counterclockwise by a single position. If you think that the clock hand is just as likely to point at 1 as it is to point at 2,

[28] The distinction between modesty and immodesty is closely related to SCHOENFIELD (2015, 2018) and STEEL (2018)'s distinction between the plans which are best to *conform to* and those which are best to *make*, or to *try* to conform to. If you are modest, then the plan which it would be best to conform to could come apart from the plan which it would be best to make. Though, if you are immodest, the plan which would be best to conform to will always be the plan which would be best to make.

3, or 4, then your credences about what you'll learn and how you'll update your credences are shown in figure 7a.

If your learning dispositions are certain to not misfire, then, at each world w in which your total evidence is e, you are certain to adopt D_e. So the accuracy of your learning dispositions at world $w \in \mathbf{T}e$ is given by $\mathscr{A}(D_e, w)$. We may then evaluate your learning dispositions with the expectation $\sum_{e \in \mathcal{E}} \sum_{w \in \mathbf{T}e} C(w) \cdot \mathscr{A}(D_e, w)$. How should we evaluate learning dispositions which may misfire? This question turns out to be a bit complicated. If your learning dispositions may misfire, then, at a world w in which your total evidence is e, it is not certain how you would update your credences, were you to adopt the dispositions D. You think there's a probability of $C(\mathbf{U}e \mid \mathbf{T}e)$ that you would correctly respond to your evidence, but for some $f \neq e$, you think there's a non-zero probability $C(\mathbf{U}f \mid \mathbf{T}e)$ that you would instead respond as if your evidence had told you f. So: if your learning dispositions may misfire, then we should say that the accuracy of your learning dispositions, at a world $w \in \mathbf{T}e$, would end up being $\mathscr{A}(D_f, w)$ with a probability of $C(\mathbf{U}f \mid \mathbf{T}e)$, for each $f \in \mathcal{E}$. So the expected accuracy of your learning dispositions at world $w \in \mathbf{T}e$ is $\sum_{f \in \mathcal{E}} C(\mathbf{U}f \mid \mathbf{T}e) \cdot \mathscr{A}(D_f, w)$. Then, we should evaluate your learning dispositions with (1).

$$\sum_{e \in \mathcal{E}} \sum_{w \in \mathbf{T}e} C(w) \cdot \sum_{f \in \mathcal{E}} C(\mathbf{U}f \mid \mathbf{T}e) \cdot \mathscr{A}(D_f, w) \tag{1}$$

Or so I think. But you may disagree. Consider the world w_1 in figure 6. This is a world in which your learning dispositions *don't* misfire. So it is certain in advance that, at w_1, the accuracy of your learning dispositions will be $\mathscr{A}(D_{\neg 3}, w_1)$. In general, for any possible world $w \in \mathbf{U}e$, you may wish to say that your accuracy at w is given by $\mathscr{A}(D_e, w)$, so that we should evaluate your learning dispositions with (2).

$$\sum_{e \in \mathcal{E}} \sum_{w \in \mathbf{U}e} C(w) \cdot \mathscr{A}(D_e, w) \tag{2}$$

I disagree because I believe that, when we are evaluating your learning dispositions, we should not ask, indicatively: how accurate *will* these learning dispositions be? Instead, we should ask, subjunctively: how accurate *would* these learning dispositions be? That is: I disagree because I am a causal decision theorist. When I evaluate a chancy act like flipping a coin at a world, w, I don't say: since the coin lands heads at w, the value of the flip at w is the value of heads. Instead, I say: *were* I to flip the coin at w, I'd have a 50% probability of heads and a 50% probability of tails, so the value of the flip at w is 50% times the value of heads plus 50% times the value of tails. Likewise: even though w_1 is a world at which I *do* update on $\neg 3$, it is not a world at which, *were* I to adopt my (potentially misfiring) learning dispositions, I *would* update on $\neg 3$. Rather, I would

have an 80% probability of updating on $\neg 3$, a 10% probability of updating on $\neg 2$, and a 10% probability of updating on $\neg 4$. So the accuracy of D at w_1 is 80% times the accuracy of $D_{\neg 3}$ plus 10% times the accuracy of $D_{\neg 2}$, plus 10% times the accuracy of $D_{\neg 4}$. For the interested reader, I have more to say about why I favor (1) in appendix A.[29]

If $\mathscr{A}$ is a nice measure of accuracy, then your (potentially misfiring) learning dispositions will maximize the expectation (2) iff you are disposed to condition on the proposition that you've updated on your total evidence. That is: you'll maximize (2) by conforming to what I'll call *update conditionalization*.[30]

> UPDATE CONDITIONALIZATION
> Be disposed to respond to total evidence e by conditioning on $\mathbf{U}e$.

(UPCONDI) $$D_e(\phi) \stackrel{!}{=} C(\phi \mid \mathbf{U}e)$$

Parenthetically: you may worry about a proposition like $\mathbf{U}e$ showing up on the right-hand-side of UPCONDI. For $\ulcorner \mathbf{U}e \urcorner$ says that you've updated on e, which means that you've taken your evidence to be e and adopted D_e in response. But the right-hand-side of UPCONDI is supposed to be telling us what D_e should be. Does this make the rule self-referential? No. $\ulcorner \mathbf{U}e \urcorner$ says only that you've taken your total evidence to be e and, in response, adopted the new credence—whatever it may be—which you are disposed to adopt in that stimulus condition. So, by including a proposition like $\mathbf{U}e$, the learning norm UPCONDI presupposes that you have some learning dispositions or other—that there is some credence you're disposed to adopt if your total evidence is e—but it does not presuppose anything about what those learning dispositions are. The norm tells you what they should be.[31]

In our simplified model of Williamson's clock, the result of updating on $\neg 4$ with UPCONDI is shown in figure 7b. After updating on the evidence that the clock hand is not at position 4, you'll think that it's most likely at position 2 (80%), though you'll save some credence for it being at 1 or 3 instead (10% each). Thus, while you'll think that your evidence likely told you that the clock hand isn't at position 4 (80%), you'll think that it could have instead told you it's not at

---

[29] Thanks to an anonymous reviewer for prompting me to say more about why I favor evaluating learning dispositions with (1) rather than (2).

[30] Since $\{\mathbf{U}e \mid e \in \mathcal{E}\}$ is a partition, this follows from Theorem 2 of GREAVES & WALLACE (2006).

[31] Of course, if you know what your learning dispositions in fact are, then you won't recognize any live possibilities in which $\mathbf{U}e$ is true but you don't adopt the credence function D_e. The point is just that the *definition* of the proposition $\ulcorner \mathbf{U}e \urcorner$ doesn't make any reference to a particular credence function, so there's no self-reference involved in the norm UPCONDI. Compare: if you know that you'll greet Rachel, then you won't recognize any live possibilities in which Rachel is not the person you greet; but this does not mean that the imperative to greet Rachel involves any self-reference. Thanks to Harvey Lederman and Adam Elga for helpful conversation on this point.

	$1 \wedge \mathbf{T}\neg 3$	$2 \wedge \mathbf{T}\neg 4$	$3 \wedge \mathbf{T}\neg 1$	$4 \wedge \mathbf{T}\neg 2$
$\mathbf{U}\neg 3$	8/40	1/40	0	1/40
$\mathbf{U}\neg 4$	1/40	8/40	1/40	0
$\mathbf{U}\neg 1$	0	1/40	8/40	1/40
$\mathbf{U}\neg 2$	1/40	0	1/40	8/40
	1/4	1/4	1/4	1/4

(a)

	$1 \wedge \mathbf{T}\neg 3$	$2 \wedge \mathbf{T}\neg 4$	$3 \wedge \mathbf{T}\neg 1$	$4 \wedge \mathbf{T}\neg 2$
$\mathbf{U}\neg 3$	0	0	0	0
$\mathbf{U}\neg 4$	1/10	8/10	1/10	0
$\mathbf{U}\neg 1$	0	0	0	0
$\mathbf{U}\neg 2$	0	0	0	0
	1/10	8/10	1/10	0

(b)

	$1 \wedge \mathbf{T}\neg 3$	$2 \wedge \mathbf{T}\neg 4$	$3 \wedge \mathbf{T}\neg 1$	$4 \wedge \mathbf{T}\neg 2$
$\mathbf{U}\neg 3$	8/100	8/100	0	0
$\mathbf{U}\neg 4$	1/100	64/100	1/100	0
$\mathbf{U}\neg 1$	0	8/100	8/100	0
$\mathbf{U}\neg 2$	1/100	0	1/100	0
	1/10	8/10	1/10	0

(c)

FIGURE 7: Given the prior credences in figure 7a, the result of conditioning on $\neg 4$ with UPCONDI is shown in figure 7b, and the result of updating on $\neg 4$ with EXCONDI is shown in figure 7c.

3 or not at 1 (10% each). So you will not be certain about what your total evidence is. So these learning dispositions permit uncertainty about what your evidence says. For externalists, this should be seen as improvement on SCONDI.

Notice that, if you are disposed to learn from your evidence in the way prescribed by UPCONDI, then you will always end up certain of how you've updated your credences. This is *prima facie* odd, given that your evidence did not tell you anything at all about your credences. By stipulation, your evidence only tells you something about the position of the clock hand. Of course, after looking at the clock, you *could* end up learning something about how your credences have changed, but you need not—perhaps you do not have introspective access to your own opinions. More generally, if rationality requires conforming to UPCONDI, then rationality forbids uncertainty about how you've updated your credences. From my perspective, this is another reason to be suspicious of evaluating learning dispositions with (2), but I wouldn't be surprised to learn that the reader disagrees.

If we instead evaluate your (potentially misfiring) learning dispositions with

the expectation (1), as I recommend, then the optimal learning dispositions will be the ones conforming to what I will call *externalist conditionalization*.

> EXTERNALIST CONDITIONALIZATION
> Be disposed to respond to the total evidence e by changing your credence in $\mathbf{T}f$ to your current credence in $\mathbf{T}f$, conditional on $\mathbf{U}e$, $C(\mathbf{T}f \mid \mathbf{U}e)$, and holding fixed your credence in each proposition conditional on $\mathbf{T}f$ (for each $f \in \mathcal{E}$).

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

(To understand why EXCONDI maximizes the expectation (1), see the proof of **Proposition 1** in appendix B.) In our simplified model of Williamson's clock, the result of updating on $\neg 4$ with EXCONDI is shown in figure 7c. When it comes to your opinions about the position of the clock hand, EXCONDI agrees with UPCONDI. You will think that the clock hand is most likely at position 2 (80%), though you'll save some credence for it being at 1 or 3 instead (10% each). For this reason, you will be uncertain about what your evidence is. Again, externalists should see this as an advantage of EXCONDI. Moreover, if you only learn about the position of the clock hand, and you don't additionally learn something about how your credences have changed, then updating with EXCONDI will leave you uncertain of how your credences have changed. You'll think that, most likely, you've updated on $\neg 4$ (66%), though you'll think that you may have updated on $\neg 3$ or $\neg 1$ instead (16% each), and you'll even put aside some credence (2%) for the possibility that you've updated on $\neg 2$.

In §2.1, we saw an argument that your learning dispositions ought to satisfy the principle of REFLECTION. In contrast to CONDI, both UPCONDI and EXCONDI will always satisfy this principle.[32] So neither of these update rules will permit the kind of deliberate self-delusion we encountered in §2.1.

Suppose that your learning dispositions are certain to not misfire—that is, suppose that you are *immodest*. Then, both UPCONDI and EXCONDI will agree

---

[32] To see that UPCONDI satisfies REFLECTION, note that the law of total probability tells us that $C(\phi) = \sum_{e \in \mathcal{E}} C(\phi \mid \mathbf{U}e) \cdot C(\mathbf{U}e)$, which, according to UPCONDI, should be $\sum_{e \in \mathcal{E}} D_e(\phi) \cdot C(\mathbf{U}e)$. To see that EXCONDI satisfies it, note that, according to EXCONDI:

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

with SCONDI.[33] As we saw above, SCONDI says to be disposed to become certain of what your evidence says. So both UPCONDI and EXCONDI say that you should be disposed to become certain of what your evidence says if you are certain in advance to not make a mistake about what your evidence says. If, however, you modestly foresee the possibility that you'll mistake your evidence, revising your opinions as if your evidence told you f, when in fact it told you e, then you should not be disposed to end up certain about what your evidence has told you.

3 Applications

The update rule EXCONDI provides a general theory of learning for the externalist. This theory tells us interesting things about the rationality of epistemic *akrasia* (§3.1), peer disagreement (§3.2), failures of *Negative Access* and undercutting defeat (§3.3), and learning without certainty (§3.4). (In this section, I will focus exclusively on EXCONDI, since this is the update rule which I think the externalist should endorse. But much of what I'll have to say here could be said with UPCONDI instead, *mutatis mutandis*.)

3.1 Epistemic Akrasia

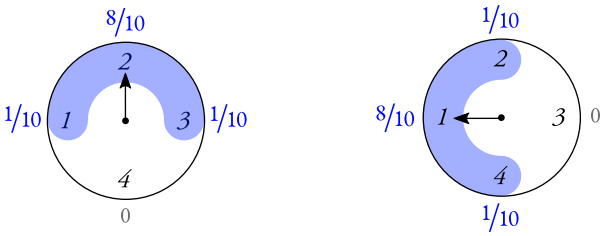
Suppose your evidence supports believing that it will rain. Since you are rational, you correctly respond to your evidence and believe that it will rain. Then, new evidence comes in. It tells you that you probably mistook some of your earlier evidence, and your belief that it will rain is likely irrational. What should you believe now? LASONEN-AARNIO (2014, 2015, forthcoming) suggests that, if your original evidence supported believing that it will rain, then, conjoined with your new evidence, it will *still* support believing that it will rain. So you should continue to believe that it will rain. Of course, your new evidence also supports believing that this is an irrational belief. So you should believe: it will rain and it's irrational to believe that it will rain. That is, you should be *episitemically akratic*.[34]

In the case of credences, LASONEN-AARNIO does not think that there are any necessary rational connections between your credences and your credences about which credences are rational. That is: she does not think that there are any *enkratic*

[33] If $C(\mathbf{U}e \leftrightarrow \mathbf{T}e) = 1$, then $C(\phi \mid \mathbf{U}e) = C(\phi \mid \mathbf{T}e)$, so UPCONDI agrees with SCONDI. And, if $C(\mathbf{U}e \leftrightarrow \mathbf{T}e) = 1$, then $C(\mathbf{T}e \mid \mathbf{U}e) = 1$, while $C(\mathbf{T}f \mid \mathbf{U}e) = 0$ if $f \neq e$. So EXCONDI says:

$$\begin{aligned} D_e(\phi) &\stackrel{!}{=} C(\phi \mid \mathbf{T}e) \cdot C(\mathbf{T}e \mid \mathbf{U}e) + \sum_{f \neq e} C(\phi \mid \mathbf{T}f) \cdot C(\mathbf{T}f \mid \mathbf{U}e) \\ &= C(\phi \mid \mathbf{T}e) \cdot 1 + \sum_{f \neq e} C(\phi \mid \mathbf{T}f) \cdot 0 \\ &= C(\phi \mid \mathbf{T}e) \end{aligned}$$

[34] See also HOROWITZ (2014) and GRECO (2014).



(a) (b)

FIGURE 8: In figure 8a, the rational credences to have, given the evidence $\neg 4$. In figure 8b, the rational credences to have, given the evidence $\neg 3$.

requirements on credences. ELGA (2013) disagrees. He believes that, for any proposition ϕ, your credence in ϕ must harmonize with your credences about which credence in ϕ is rational. Return to our simplified model of Williamson's clock. Suppose your evidence tells you that the clock hand is not at position 4, you update on this evidence with EXCONDI, and arrive at the credences $D_{\neg 4}$ (figure 8a). These are the rational credences to hold, given your evidence, but you're not certain about whether they are rational. You think the clock hand might be at position 1, in which case $D_{\neg 3}$ (figure 8b) are the rational credences to have. Conditional on $D_{\neg 3}$ being the rational credences, which credences should you have? A natural first thought is this: conditional on $D_{\neg 3}$ being rational, your credences should agree with $D_{\neg 3}$. In general, conditional on D_f being rational, your credences should agree with D_f. Call this principle *rational reflection*.[35]

> RATIONAL REFLECTION
> Conditional on D_f being the rational credences for you to hold, your credences should agree with D_f.

$$D_e(\phi \mid D_f \text{ is rational}) = D_f(\phi)$$

(RAT REF) $$D_e(\phi \mid \mathbf{T}f) = D_f(\phi)$$

(Here, I'm assuming that your evidence was *actually* e, so that D_e are actually the rational credences for you to hold. And I'm assuming that you're certain that D_f is rational iff your total evidence was f, $\mathbf{T}f$.)

ELGA thinks that RAT REF is not quite right. If $D_{\neg 3}$ is rational, then your evidence tells you that the clock hand isn't at 3. And you are certain that your evidence tells you this iff the clock hand is at 1. So, conditional on $D_{\neg 3}$ being rational, you should be *certain* that the clock hand points at 1. But $D_{\neg 3}$ is not certain of that. The reason is that $D_{\neg 3}$ is not certain that it is rational. So, conditional on $D_{\neg 3}$ being rational, you are assured of something which $D_{\neg 3}$ is not. Before

[35] See CHRISTENSEN (2010)

you align your credences with $D_{\neg 3}$, you should assure it that it is rational.[36] So what's *exactly* right isn't RAT REF, but instead the principle ELGA calls *new rational reflection*.

> NEW RATIONAL REFLECTION
> Conditional on D_f being the rational credences for you to hold, your credences should agree with D_f, once D_f is informed that it is rational.

$$D_e(\phi \mid D_f \text{ is rational}) = D_f(\phi \mid D_f \text{ is rational})$$

(NEW RAT REF) $$D_e(\phi \mid \mathbf{T}f) = D_f(\phi \mid \mathbf{T}f)$$

This is an *enkratic* principle which says how your views about the requirements of rationality should constrain your other views. It is the kind of principle which LASONEN-AARNIO rejects. Indeed, LASONEN-AARNIO (2015) argues against NEW RAT REF by showing that, if an externalist learns from their evidence in the way prescribed by CONDI, then they will sometimes violate the principle. To my mind, this is not a reason to reject NEW RAT REF, but rather yet another reason why an externalist should reject CONDI. Notice that, if an externalist updates their credences in accordance with EXCONDI, they will always satisfy NEW RAT REF.[37] So if you abide EXCONDI, you will always be epistemically enkratic. When you are uncertain about whether you are rational, you will see reason to think that your credences are irrational as a reason to revise those credences.

(Note: there are two ways to be uncertain about what rationality requires of you. You could be certain about what rationality requires if your evidence tells you e, what it requires if your evidence tells you f, and so on, but be uncertain what your evidence tells you. This is the kind of uncertainty about the requirements of rationality which arises from externalism. Alternatively, you could be uncertain about what rationality requires of you when your evidence tells you e. That is: you could be uncertain about the *a priori* requirements of rationality. LASONEN-AARNIO thinks that *both* kinds of uncertainty can be rational. For an

---

[36] *Cf.* HALL (1994) and LEWIS (1994).

[37] To see this, note that, for any $e, f \in \mathcal{E}$, $D_e(\phi \mid \mathbf{T}f) = C(\phi \mid \mathbf{T}f)$.

$$\begin{aligned} D_e(\phi \mid \mathbf{T}f) &= \frac{D_e(\phi \wedge \mathbf{T}f)}{D_e(\mathbf{T}f)} \\ &= \frac{\sum_{g \in \mathcal{E}} C(\phi \wedge \mathbf{T}f \mid \mathbf{T}g) \cdot C(\mathbf{T}g \mid \mathbf{U}e)}{\sum_{g \in \mathcal{E}} C(\mathbf{T}f \mid \mathbf{T}g) \cdot C(\mathbf{T}g \mid \mathbf{U}e)} \\ &= \frac{C(\phi \mid \mathbf{T}f) \cdot C(\mathbf{T}f \mid \mathbf{U}e)}{C(\mathbf{T}f \mid \mathbf{T}f) \cdot C(\mathbf{T}f \mid \mathbf{U}e)} \\ &= C(\phi \mid \mathbf{T}f) \end{aligned}$$

So $D_e(\phi \mid \mathbf{T}f)$ will equal $D_f(\phi \mid \mathbf{T}f)$, since they are both equal to $C(\phi \mid \mathbf{T}f)$.

	$1 \wedge \mathbf{T}\neg 3$	$2 \wedge \mathbf{T}\neg 4$	$3 \wedge \mathbf{T}\neg 1$	$4 \wedge \mathbf{T}\neg 2$
$\mathbf{U}\neg 3$	8/20	1/20	0	0
$\mathbf{U}\neg 4$	1/20	8/20	0	0
$\mathbf{U}\neg 1$	0	1/20	0	0
$\mathbf{U}\neg 2$	1/20	0	0	0
	1/2	1/2	0	0

FIGURE 9: If your epistemic peer is just as likely to respond correctly to their evidence as you are, and how they respond to their evidence is independent of how you do, then this is the result of updating the credence $D_{\neg 4}$ from figure 7c on the new evidence that your peer updated on $\neg 3$.

argument that the second kind of uncertainty is never rationally permitted, see TITELBAUM (2015). Throughout, my discussion is restricted to only the first kind of uncertainty.)

3.2 PEER DISAGREEMENT

Suppose it wasn't just you that looked at the clock. Suppose an *epistemic peer* of yours also took a glimpse at the clock, in exactly the same circumstances as you, so that you are certain that both you and your peer learned the same thing. Suppose that, like you, your peer is 80% likely to respond rationally to their evidence, but they have a 20% probability of mistaking their evidence, and they are equally likely to err in either the clockwise or counterclockwise direction. For the sake of concreteness, suppose you think that whether/how your learning dispositions misfire is independent of whether/how your peer's learning dispositions misfire. Both you and your peer learn $\neg 4$. You rationally respond to your evidence and end up with the credences shown in figures 7c and 8a. Your peer, on the other hand, responds *irrationally*, and ends up thinking that that clock hand is 80% likely to be at position 1 (as in figure 8b).

Applied to this case, the question of peer disagreement is whether learning about your peer's opinion gives you any reason to revise your views about the position of the clock hand. In general, a *conciliationist* says: learning that an epistemic peer disagrees with you *does* give you reason to revise your views. In cases like this, EXCONDI agrees with the conciliationist. Suppose you have the credences in figure 7c, and then learn how your peer updated *their* credences. When you learn this, your total evidence will be either that your peer has updated on $\neg 1$, that they've updated on $\neg 2$, that they've updated on $\neg 3$, or that they've updated on $\neg 4$. Suppose, for simplicity's sake, that you are immodestly certain to respond correctly to this evidence (though, of course, you still modestly think you may have responded incorrectly to your evidence about the clock hand's position). Then, if you learn that your peer updated on $\neg 3$, you will end up with the credences shown in figure 9. The disagreement of your peer has given you reason to

revise your views about where the clock hand is located. You now think that it's 50% likely to be at 1 and 50% likely to be at 2. (Since your peer was certain to not update on $\neg 3$ if the clock hand were pointing at 3, you are now certain that it's not at 3.)

More generally, if you are modest and you learn from your evidence in the way EXCONDI says you should, then you can end up thinking that your credence in ϕ might be irrational, and you will satisfy NEW RAT REF, so you will see reason to think that your credence in ϕ is irrational as reason to change that credence. In general, learning that an epistemic peer has responded differently to your shared evidence can give you reason to think that your learning dispositions have misfired, and that your credence that ϕ is irrational. And so: learning that an epistemic peer has responded differently to your shared evidence can give you reason to revise your credence in ϕ.[38]

3.3 Failures of Negative Access and Undercutting Defeat

Thus far, I have been focusing on cases, like Williamson's clock, in which *Positive Access* is violated. But EXCONDI applies in cases in which *Negative Access* is violated as well. Suppose you are about to bite into a pear and learn whether it tastes sweet. Before biting, you think it's 50% likely to taste sweet. Incidentally, you also think it's 20% likely that Sabeen has slipped you a psychotropic drug. This drug renders you incapable of properly categorizing flavor experiences. Drug-free, you are able to recognize how things taste to you. Drugged, your opinions about how things taste correlate not at all with the way they actually taste to you. When drugged, you are not able to tell that you are drugged. You bite into the pear. In the good case, you are drug free, and your evidence tells you that the pear tastes sweet to you. In the bad case, you are drugged, and you have no evidence about how the pear tastes to you, though you end up unreasonably confident that it tastes sweet. In the bad case, your evidence doesn't tell you that the pear is sweet, $\neg\mathbf{E}s$, nor does your evidence tell you that you're not in the good case; so your evidence doesn't say that it doesn't say that the pear is sweet, $\neg\mathbf{E}\neg\mathbf{E}s$. So *Negative Access* is violated.

Suppose that, if you're drug-free, $\neg d$, then you will learn how the pear tastes to you. If it tastes sweet, s, then you will learn that it tastes sweet, $\mathbf{T}s$. If it doesn't taste sweet, $\neg s$, then you will learn that it doesn't taste sweet, $\mathbf{T}\neg s$. If you are drugged, d, then you won't learn anything at all, $\mathbf{T}\top$. If you are drug-free, then you will update on whatever your evidence tells you. However, if you are drugged, then your learning dispositions will mis-fire, and you'll incorrectly update on either s or $\neg s$ (with equal probability), even though you've learned neither. Before biting into the pear, you have the credences shown in figure 10a. In fact, you are drug-free, and the pear tastes sweet. If you update on s with

[38] See STEEL (2018) and SCHOENFIELD (2018) for similar justifications of conciliationism.

	$\neg d \wedge s \wedge \mathbf{T}s$	$\neg d \wedge \neg s \wedge \mathbf{T}\neg s$	$d \wedge s \wedge \mathbf{T}\top$	$d \wedge \neg s \wedge \mathbf{T}\top$
$\mathbf{U}s$	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) Before tasting the pear

	$\neg d \wedge s \wedge \mathbf{T}s$	$\neg d \wedge \neg s \wedge \mathbf{T}\neg s$	$d \wedge s \wedge \mathbf{T}\top$	$d \wedge \neg s \wedge \mathbf{T}\top$
$\mathbf{U}s$	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) The result of updating on s with EXCONDI

	$\neg d \wedge s \wedge \mathbf{T}s$	$\neg d \wedge \neg s \wedge \mathbf{T}\neg s$	$d \wedge s \wedge \mathbf{T}\top$	$d \wedge \neg s \wedge \mathbf{T}\top$
$\mathbf{U}s$	8/10	0	1/10	0
$\mathbf{U}\neg s$	0	0	1/10	0
	8/10	0	2/10	0

(c) The result of updating on s with CONDI

FIGURE 10

EXCONDI, you'll arrive at the new credences shown in figure 10b. You'll continue thinking that it's 20% likely you've been drugged, and your credence that the pear is sweet will rise from 50% to 90%.

Contrast EXCONDI's advice about this case with the advice of CONDI. If you abide CONDI, then, after learning that the pear tastes sweet, you will be *certain* that the pear tastes sweet (figure 10c). Conditioning can never lower a proposition's credence from one, so once you are certain that the pear tastes sweet, you will remain certain that it tastes sweet forever after. But suppose you receive (misleading) evidence that Sabeen drugged you. Evidence like this provides *undercutting defeat* for your high credence that the pear tastes sweet. If you've reason to raise your credence that d, then you've reason to *lower* your credence that s. Nonetheless, if you're disposed to learn from your evidence in the way CONDI recommends, then your credence in s will not be affected by learning that you're likely drugged.[39] Not so with EXCONDI. Updating on s with EXCONDI introduces a dependence between your credence that d and your credence that s. Once you've learned from your evidence in the way recommended by EXCONDI, you'll see misleading evidence that you were drugged as a reason to become *less* confident that the pear tastes sweet.[40]

[39] See CHRISTENSEN (1992), WEISBERG (2009, 2015), and GALLOW (2014).

[40] A defender of CONDI may suggest that, instead of conditioning on s, you should instead condition on the material conditional $\neg d \rightarrow s$ (WAGNER, 2013). This will have the unfortunate consequence of raising your credence that you've been drugged from 20% to over 33% (see GALLOW, 2014). A defender of CONDI could also, of course, just deny that there are failures of

In my younger and more vulnerable years, I suggested an update rule, called *holistic conditionalization*, for learning episodes like these (GALLOW, 2014). In this case, the input to the rule would be the set of ordered pairs $\{< \neg d, s >, < d, \top >\}$, which I gave the following interpretation: if $\neg d$ is true, then your total evidence is s; and, if d is true, you have no evidence at all. More generally, holistic conditionalization says how to be disposed to revise your credences given a set of ordered pairs $\{< t_i, e_i >\}_i$, with the interpretation that, for each i, if t_i is true, then your total evidence is e_i.

> HOLISTIC CONDITIONALIZATION
> Given the input $E = \{< t_i, e_i >\}_i$, be disposed to adopt the new credence function D_E, where, for each proposition ϕ,
>
> (HCONDI) $$D_E(\phi) = \sum_i C(\phi \mid t_i \wedge e_i) \cdot C(t_i)$$

In this case, HCONDI and EXCONDI agree. If you follow HCONDI's advice about how to revise the credences from figure 10a, given the input $\{< \neg d, s >, < d, \top >\}$, then you will get the same result as updating on s with EXCONDI (figure 10b).

More generally, suppose that, if the background theory t is true, then you will learn the true member of $\{e_1, e_2, \ldots, e_N\}$, where these potential evidence propositions are mutually exclusive and jointly exhaustive. If t is false, then you'll not learn anything at all, though you'll still update as though you had learned one of the propositions in $\{e_1, e_2, \ldots, e_N\}$. This is the kind of 'theory-dependent' experiment which HCONDI was designed to handle. And, in this experiment, the result of updating on the evidence e with EXCONDI will be:[41]

$$D_e(\phi) = C(\phi \mid t \wedge e) \cdot C(t \mid \mathbf{U}e) \; + \; C(\phi \mid \neg t) \cdot C(\neg t \mid \mathbf{U}e)$$

If whether you update on e is independent of whether the background theory t is true, $C(t \mid \mathbf{U}e) = C(t)$, then EXCONDI will deliver exactly the same result as updating on the input $\{< t, e >, < \neg t, \top >\}$ with HCONDI. If $\mathbf{U}e$ is not independent of t, then EXCONDI and HCONDI need not agree. But this is for the good, since the fact that HCONDI always holds fixed your credence in the background theory t is a problem for that rule. Suppose you know that the chance of the pear tasting sweet is one in a million, while the chance of Sabeen slipping you the drug is one half. Then, you should be disposed to become very confident that you've been slipped the drug if you learn that the pear tastes sweet. But HCONDI disagrees, saying that you should remain 50% confident that you've been slipped the

Negative Access like this. If externalism is false, then CONDI won't face objections like these.

[41] To see this, note that: 1) you are certain in advance that $\mathbf{T}e \leftrightarrow t \wedge e$, for each e; 2) you are certain that $\mathbf{T}\top \leftrightarrow \neg t$; and 3) if t is true, you will update on e iff e is true, so that $C(t \wedge e \mid \mathbf{U}e) = C(t \mid \mathbf{U}e)$ and $C(t \wedge e \mid \mathbf{U}f) = 0$ if $e \neq f$.

	$b \wedge \mathbf{T}b$	$v \wedge \mathbf{T}v$	$g \wedge \mathbf{T}g$
$\mathbf{U}b$	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) Before looking

	$b \wedge \mathbf{T}b$	$v \wedge \mathbf{T}v$	$g \wedge \mathbf{T}g$
$\mathbf{U}b$	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) The result of updating on b with EXCONDI

FIGURE 11

drug.[42] In contrast, EXCONDI says that, if $C(d \mid \mathbf{U}s) \gg C(d)$, then you should be disposed to become much more confident that you've been slipped the drug upon learning that the pear tastes sweet.

3.4 Learning without Certainty

Thus far, I have been focusing on cases in which it becomes rational for you to become *certain* that some proposition is true. But EXCONDI does not require certainty in any new proposition. Suppose that you are about to observe a cloth in dim lighting. You know the cloth is either blue, violet, or green, and you think each color is equally likely. You know that your experience will teach you its true color. But, because the lighting is so dim, you will find your experience difficult to discern, and so 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 take it to be violet (20%) or green (10%) instead. Similarly, if it's violet, you'll most likely recognize it as violet (70%), though you may incorrectly think it's blue (20%) or green (10%). And, if it's green, you'll most likely recognize it as green (80%), though perhaps you'll instead mistake it for blue or violet (10% each). Then, before looking at the cloth, your credences are as shown in figure 11a. Suppose you look at the cloth and learn that it is blue. EXCONDI says that you should be disposed to respond to this evidence by becoming 70% confident that the cloth is blue, 20% confident that it is violent, and 10% confident that it is green. The result of updating the distribution from figure 11a on b with EXCONDI is shown in figure 11b. Notice that, though your credences about the cloth's color have changed, you have not become certain of any proposition.

[42] I recognized this problem and suggested a rule for updating your credences in the background theories, but the rule was overly complicated and under-motivated; I no longer accept it.

Learning episodes like this are discussed by JEFFREY (1965) (though JEFFREY thinks about them differently from the way I am suggesting we think about them here).[43] In JEFFREY's treatment, the input to a revision of your credences is a set of ordered pairs $\{< e_i, \alpha_i >\}_i$ of propositions, e_i, and real numbers, α_i—and JEFFREY supposes that the e_i are mutually exclusive and jointly exhaustive, and that the α_i sum to 1. Inputs like these are eponymously called 'Jeffrey shifts'. The interpretation of a Jeffrey shift like this is that α_i is the new credence in e_i which has been rationalized by experience. In the case of the dimly-lit cloth, your Jeffrey shift may be $\{< b, 70\% >, < v, 20\% >, < g, 10\% >\}$. If so, then the result of updating on b with EXCONDI (figure 11b) is exactly what is recommended by JEFFREY's learning norm, known as:

> JEFFREY CONDITIONALIZATION
> Given the input $E = \{< e_i, \alpha_i >\}_i$, be disposed to adopt the new credence function D_E, where, for each proposition ϕ,
>
> (JCONDI) $$D_E(\phi) = \sum_i C(\phi \mid e_i) \cdot \alpha_i$$

More generally, suppose that experience will teach the true member of $\{e_1, e_2, \ldots, e_N\}$, where these e_i are mutually exclusive and jointly exhaustive. Then, updating with EXCONDI on e_j is equivalent to updating with JCONDI on the Jeffrey shift $\{< e_i, C(\mathbf{T}e_i \mid \mathbf{U}e_j) >\}_i$.[44]

To repeat: the way that JEFFREY thought about Jeffrey shifts is quite different from the way I am suggesting we think about them. However, it is still noteworthy that EXCONDI provides us with a way of understanding learning episodes which don't rationalize certainty in any proposition; and that, equipped with this understanding, the advice of EXCONDI aligns with the advice of JCONDI, given a Jeffrey shift over the natural propositions.

4 IN SUMMATION

My goal has been to say how you should be disposed to revise your opinions in light of your evidence, if externalism is true. Co-opting an argument from SALOW (2018), I've contended that, if externalism is true, then we should reject the orthodox Bayesian learning norm conditionalization. If externalism is true,

---

[43] See also FIELD (1978), who presents a rule similar to JEFFREY's, and suggests a different way of thinking about these learning experiences.

[44] Since experience will teach exactly one e_i, $\{\mathbf{T}e_1, \mathbf{T}e_2, \ldots, \mathbf{T}e_N\}$ is a partition. Since experience will teach the *true* e_i, $\mathbf{T}e_i$ entails e_i. Since the e_i are mutually exclusive and jointly exhaustive, it follows that $\mathbf{T}e_i \leftrightarrow e_i$ is certain, so that $C(\phi \mid \mathbf{T}e_i) = C(\phi \mid e_i)$. Then, EXCONDI says $D_{e_j}(\phi) \overset{!}{=} \sum_i C(\phi \mid \mathbf{T}e_i) \cdot C(\mathbf{T}e_i \mid \mathbf{U}e_j) = \sum_i C(\phi \mid e_i) \cdot C(\mathbf{T}e_i \mid \mathbf{U}e_j)$, which is the result of updating with JCONDI on $\{< e_i, C(\mathbf{T}e_i \mid \mathbf{U}e_j) >\}_i$.

then conditionalization allows you to engage in deliberate self-delusion, disposing yourself to become more confident of some proposition, so long as that proposition is false. And learning dispositions like these are not rational. To seek out a replacement for conditionalization, I supposed that learning dispositions which maximize expected accuracy are rational. And I argued that, if your dispositions to learn from your evidence are not perfect—if you modestly foresee some possibility of mistaking your evidence—then the expected accuracy maximizing learning dispositions are those conforming to externalist conditionalization. Learning dispositions like these will never permit the kind of deliberate self-delusion which conditionalization condoned. This theory of rational learning allows uncertainty about what your evidence is; and so, it permits uncertainty about whether your credences are rational or not. It forbids a form of epistemic *akrasia*, counsels conciliation in certain cases of peer disagreement, appropriately handles cases of undercutting defeat, and permits learning without certainty.

A Evaluating Learning Dispositions

In this appendix, I will say a bit more about why I think we should evaluate potentially misfiring learning dispositions with the expectation (1).

Causal decision theorists say that you should evaluate an act, A, with[45]

$$\sum_{w^*} C_A(w^*) \cdot \mathcal{V}(w^*) \tag{3}$$

where $\mathcal{V}(w^*)$ is the value of the world w^* and C_A is your credence function *imaged* on A. As I will understand it, C_A comes from your views about how likely it is that various possibilities *would* result, *were* you to perform the act A. To capture these views, we'll introduce a probability function, w_A, for each world w. If you think it's x likely that world w^* would result, were you to perform A in w, then let's say that $w_A(w^*) = x$.[46] From these probability functions, w_A, and your credences, we define C_A. We use $w_A(w^*)$ to see what proportion of your credence in w should be transferred to the world w^*. C_A is the result of carrying out these transfers for every possible world w. That is: to get the credence C_A gives to w^*, you sum up a proportion $w_A(w^*)$ of the credence C gives to w, for each w.

$$C_A(w^*) = \sum_{w} w_A(w^*) \cdot C(w) \tag{4}$$

Putting together (3) and (4) tells us that the causal decision theorist evaluates acts with

$$\sum_{w^*} \sum_{w} w_A(w^*) \cdot C(w) \cdot \mathcal{V}(w^*) \tag{5}$$

Greaves (2013) raises a worry about evaluating learning dispositions with causal decision theory. The worry is that causal decision theory advises you to accept 'epistemic bribes', sacrificing accuracy in one proposition so as to cause yourself to have greater accuracy in others. I've been persuaded by Konek & Levinstein (2019) that the proper response to these kinds of worries is to re-conceptualize what causal decision theory is saying, and distinguish between *practical* and *epistemic* value. They note that, if we define

$$\mathcal{V}_A(w) \stackrel{\text{def}}{=} \sum_{w^*} w_A(w^*) \cdot \mathcal{V}(w^*) \tag{6}$$

---

[45] See, for instance, Lewis (1981) and Joyce (1999). Not all formulations of causal decision theorists utilize *imaging* functions in this way; but I'll focus on these formulations here.

[46] It's important that we interpret the probability functions w_A in this way. Suppose, instead, we set $w_A(w^*) = C(w^* \mid A)$. Then, causal decision theory would reduce to *evidential* decision theory, since

$$C_A(w^*) = \sum_{w} w_A(w^*) \cdot C(w) = \sum_{w} C(w^* \mid A) \cdot C(w) = C(w^* \mid A) \cdot \sum_{w} C(w) = C(w^* \mid A)$$

In which case, imaging C on A reduces to conditioning C on A.

Then (5) may be re-written as:

$$\sum_{w} C(w) \cdot \mathcal{V}_A(w) \tag{7}$$

The interpretation is that $\mathcal{V}_A(w)$ tells us how practically valuable the act A is at the world w. Corresponding to this algebraic trick is a shift in perspective: we don't see causal decision theory as saying that you should use *subjunctive beliefs* to evaluate acts. Instead, it says something about how acts are to be evaluated. It says: you should measure the value of an act at a world by considering the value of what *would* result from the act's performance at that world.

According to (6), the practical value of A at w is a function of two kinds of inputs: 1) probabilities, $w_A(w^*)$, which tell you how likely it is that performing A in w would bring about w^*; and 2) values, $\mathcal{V}(w^*)$, which tell you how valuable the world w^* is. The probabilities $w_A(w^*)$ are necessary because, in some cases, it may not be certain what would result from A's performance in w. Using the values $\mathcal{V}(w^*)$ in (6) (rather than $\mathcal{V}(w)$, say) encodes the causalist's committment to value acts in terms of the good they are able to *bring about*—the good they are in a position to *causally promote*.

In the epistemic case, I think that we should evaluate learning dispositions in essentially the same way as (7) evaluates acts. That is: we should use an expectation of the form

$$\sum_{w} C(w) \cdot \mathcal{V}_D(w) \tag{8}$$

where $\mathcal{V}_D(w)$ says how epistemically valuable the learning dispositions D are at the world w. And I think that we should value the learning dispositions D, at w, with:

$$\mathcal{V}_D(w) = \sum_{w^*} w_D(w^*) \cdot \mathscr{A}(D_{w^*}, w) \tag{9}$$

('D_{w^*}' are the credences you adopt in w^*.) Here, as before, '$w_D(w^*)$' tells us how likely it is that adopting D would bring about w^*. As before, this is necessary because, when your learning dispositions may misfire, it will not be certain what would result from adopting the learning dispositions D at w. Now, if epistemic value were just like practical value, I would have written '$\mathscr{A}(D_{w^*}, w^*)$', instead of '$\mathscr{A}(D_{w^*}, w)$'. That would amount to saying: what's valuable in learning dispositions is the accuracy that they are able to *bring about*—the accuracy that they are in a position to *causally promote*. But I think, with Konek & Levinstein, that this gets the direction-of-fit of doxastic states wrong. Doxastic states don't properly aim to change the world, but rather to accurately reflect it. So when we evaluate the learning dispositions D at w, we should ignore whatever accuracy those learning dispositions are in a position to causally promote; we shouldn't think about how accurate D would *make themselves*. We should only think about how accurately they would reflect w. So I think that (9) gives the epistemic value of D at w.

Pick an arbitrary w, and let e be your total evidence in w. Then, we should suppose that, for every $f \in \mathcal{E}$, w_D gives a probability of $C(\mathbf{U}f \mid \mathbf{T}e)$ to the proposition $\mathbf{U}f$. That is, if $w \in \mathbf{T}e$, then $w(\mathbf{U}f) = C(\mathbf{U}f \mid \mathbf{T}e)$. For you think that, when your total evidence is e, there's a probability of $C(\mathbf{U}f \mid \mathbf{T}e)$ that your learning dispositions would lead you

to update on f. And the function w_D should reflect this.[47]

Then, the epistemic value of D at a world $w \in \mathbf{T}e$, will be:

$$\begin{aligned}
\mathcal{V}_D(w) &= \sum_{f \in \mathcal{E}} \sum_{w^* \in \mathbf{U}f} w_D(w^*) \cdot \mathscr{A}(D_f, w) \\
&= \sum_{f \in \mathcal{E}} \mathscr{A}(D_f, w) \cdot \sum_{w^* \in \mathbf{U}f} w_D(w^*) \\
&= \sum_{f \in \mathcal{E}} \mathscr{A}(D_f, w) \cdot w_D(\mathbf{U}f) \\
&= \sum_{f \in \mathcal{E}} \mathscr{A}(D_f, w) \cdot C(\mathbf{U}f \mid \mathbf{T}e) \qquad (10)
\end{aligned}$$

And putting (10) together with (8) gives us the expectation (1).

B Technicalities

Proposition 1. *Given any strictly proper, additive, and extensional measure of accuracy, the potentially misfiring learning dispositions which maximize the expectation* (1) *are the ones conforming to* EXCONDI.

Proof. If $\mathscr{A}$ is additive, then $\mathscr{A}(D_e, w)$ has the form $\sum_\phi \lambda_\phi \cdot \mathscr{A}(D_e(\phi), \phi, w)$, for some weights $\lambda_\phi > 0$ and some function $\mathscr{A}(x, \phi, w)$ of the accuracy of a credence x in the proposition ϕ in a world w. So (1) is:

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

Pick a ϕ and pick an $e \in \mathcal{E}$. Let $x \stackrel{\text{def}}{=} D_e(\phi)$. 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

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

---

[47] As with the practical case, it's important that we interpret the probability functions w_D correctly. (See footnote 46 above.)

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

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

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

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

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

$\sum_{f\in\mathcal{E}} C(- \mid \mathbf{T}f)\cdot C(\mathbf{T}f \mid \mathbf{U}e)$ 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 $D_e(\phi)$ which maximizes (1) is:

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

□

References

Berker, Selim. 2013. "Epistemic Teleology and the Separateness of Propositions." *Philosophical Review*, vol. 122 (3): 337–393. [14]

Blackwell, Kevin & Daniel Drucker. 2019. "When Propriety is Improper." *Philosophical Studies*, vol. 176 (2): 367–386.

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

Briggs, R. A. & Richard Pettigrew. forthcoming. "An Accuracy-Dominance Argument for Conditionalization." *Noûs*. [2]

Bronfman, Aaron. 2014. "Conditionalization and not Knowing that One Knows." *Erkenntnis*, vol. 79 (4): 871–892. [14]

Caie, Michael. 2013. "Rational Probabilistic Incoherence." *Philosophical Review*, vol. 122 (4): 527–575.

—. 2018. "A Problem for Credal Consequentialism." In *Epistemic Consequentialism*, Jeffrey Dunn & Kristoffer Ahlstrom-Vig, editors. Oxford University Press, Oxford.

Carr, Jennifer. 2017. "Epistemic Utility Theory and the Aim of Belief." *Philosophy and Phenomenological Research*, vol. 95 (3): 511–534. [14]

Christensen, David. 1992. "Confirmational Holism and Bayesian Epistemology." *Philosophy of Science*, vol. 59 (4): 540–557. [26]

—. 2010. "Rational Reflection." *Philosophical Perspectives*, vol. 24 (1): 121–140. [23]

Easwaran, Kenny. 2013. "Expected Accuracy Supports Conditionalization—and Conglomerability and Reflection." *Philosophy of Science*, vol. 80 (1): 119–142. [14]

Elga, Adam. 2013. "The puzzle of the unmarked clock and the new rational reflection principle." *Philosophical Studies*, vol. 164 (1): 127–139. [2], [22], [23]

Field, Hartry. 1978. "A Note on Jeffrey Conditionalization." *Philosophy of Science*, vol. 45 (3): 361–367. [29]

Fitelson, Branden, Kenny Easwaran & David McCarthy. ms. *Coherence*. [14]

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. [3], [26], [27]

—. 2017. "Diachronic Dutch Books and Evidential Import." *Philosophy and Phenomenological Research*, vol. 99 (1): 49–80. [2]

—. 2019. "Learning and Value Change." *Philosophers' Imprint*, vol. 19 (29): 1–22. [2]

Greaves, Hilary. 2013. "Epistemic Utility Theory." *Mind*, vol. 122 (488): 915–952. [14], [31]

Greaves, Hilary & David Wallace. 2006. "Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility." *Mind*, vol. 115 (495): 607–632. [2], [7], [14], [18]

Greco, Daniel. 2014. "A puzzle about epistemic akrasia." *Philosophical Studies*, vol. 167 (2): 201–219. [2], [22]

Hall, Ned. 1994. "Correcting the Guide to Objective Chance." *Mind*, vol. 103 (412): 505–517. [23]

Hild, Matthias. 1998a. "Auto-Epistemology and Updating." *Philosophical Studies*, vol. 92 (3): 321–361. [11], [14]

—. 1998b. "The Coherence Argument Against Conditionalization." *Synthese*, vol. 115 (2): 229–258. [11], [14]

Horowitz, Sophie. 2014. "Epistemic Akrasia." *Noûs*, vol. 48 (4): 718–744. [2], [22]

Jeffrey, Richard. 1965. *The Logic of Decision*. McGraw-Hill, New York. [3], [29], [30]

Joyce, James M. 1998. "A Nonpragmatic Vindication of Probabilism." *Philosophy of Science*, vol. 65 (4): 575–603. [13], [14]

—. 1999. *The Foundations of Causal Decision Theory*. Cambridge University Press, Cambridge. [31]

—. 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. [13], [14]

Konek, Jason & Benjamin A Levinstein. 2019. "The Foundations of Epistemic Decision Theory." *Mind*, vol. 128 (509): 69–107. [31], [32]

Lange, Marc. 1999. "Calibration and the Epistemological Role of Bayesian Conditionalization." *Journal of Philosophy*, vol. 96 (6): 294–324. [2]

Lasonen-Aarnio, Maria. 2014. "Higher Order Evidence and the Limits of Defeat." *Philosophy and Phenomenological Research*, vol. 88 (2): 314–345. [2], [22]

—. 2015. "New Rational Reflection and Internalism about Rationality." In *Oxford Studies in Epistemology*, Tamar Szabó Gendler & John Hawthorne, editors, vol. 5, chap. 5. Oxford University Press, Oxford. [2], [22], [24]

—. forthcoming. "Enkrasia or Evidentialism? Learning to Love Mismatch." *Philosophical Studies*. [2], [22]

Leitgeb, Hannes & Richard Pettigrew. 2010a. "An Objective Justification of Bayesianism I: Measuring Inaccuracy." *Philosophy of Science*, vol. 77 (2): 201–235. [14]

—. 2010b. "An Objective Justification of Bayesianism II: The Consequences of Minimizing Inaccuracy." *Philosophy of Science*, vol. 77 (2): 236–272. [2], [14]

Levinstein, Benjamin Anders. 2012. "Leitgeb and Pettigrew on Accuracy and Updating." *Philosophy of Science*, vol. 79 (3): 413–424. [14]

Lewis, David K. 1981. "Causal Decision Theory." *Australasian Journal of Philosophy*, vol. 59 (1): 5–30. [31]

—. 1994. "Humean Supervenience Debugged." *Mind*, vol. 103 (412): 473–490. [23]

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

Oddie, Graham. 2019. "What Accuracy Could Not Be." *The British Journal for the Philosophy of Science*, vol. 70 (2): 551–580.

Pettigrew, Richard. 2011. "An Improper Introduction to Epistemic Utility Theory." In *EPSA Philosophy of Science: Amsterdam 2009*, Henk W. de Regt, Stephan Hartmann & Samir Okasha, editors. Springer. [14]

—. 2012. "Accuracy, Chance, and the Principal Principle." *Philosophical Review*, vol. 121 (2): 241–275. [14]

—. 2016a. *Accuracy and the Laws of Credence*. Oxford University Press, Oxford. [13], [14]

—. 2016b. "Accuracy, Risk, and the Principle of Indifference." *Philosophy and Phenomenological Research*, vol. 92 (1): 35–59. [14]

—. 2018. "Making things right: the true consequences of decision theory in epistemology." In *Epistemic Consequentialism*, Jeffrey Dunn & Kristoffer Ahlstrom-Vig, editors. Oxford University Press, Oxford. [14]

Pryor, James. 2004. "What's Wrong with Moore's Argument?" *Philosophical Issues*, vol. 14 (1): 349–378. [10]

Salow, Bernhard. 2018. "The Externalist's Guide to Fishing for Compliments." *Mind*, vol. 127 (507): 691–728. [2], [6], [9], [11], [13], [30]

Schoenfield, Miriam. 2015. "Bridging Rationality and Accuracy." *Journal of Philosophy*, vol. 112 (12): 633–657. [14], [16]

—. 2017a. "Conditionalization Does Not (in General) Maximize Expected Accuracy." *Mind*, vol. 126 (504): 1155–1187.

—. 2017b. "The Accuracy and Rationality of Imprecise Credences." *Noûs*, vol. 51 (4): 667–685. [14]

—. 2018. "An Accuracy Based Approach to Higher Order Evidence." *Philosophy and Phenomenological Research*, vol. 96 (3): 690–715. [14], [16], [25]

Stalnaker, Robert C. 2009. "On Hawthorne and Magidor on Assertion, Context, and Epistemic Accessibility." *Mind*, vol. 118 (470): 399–409. [6]

Steel, Robert. 2018. "Anticipating Failure and Avoiding It." *Philosophers' Imprint*, vol. 18 (13): 1–28. [16], [25]

Teller, Paul. 1976. "Conditionalization, Observation, and Change of Preference." In *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, W. L. Harper & C. A. Hooker, editors, vol. I, 205–253. D. Reidel Publishing Company, Dordrecht. [2]

Titelbaum, Michael G. 2013. *Quitting Certainties: A Bayesian Framework Modeling Degrees of Belief*. Oxford University Press, Oxford. [2]

—. 2015. "Rationality's Fixed Point (Or: In Defense of Right Reason)." *Oxford Studies in Epistemology*, vol. 5: 253–294. [24]

van Fraassen, Bas C. 1984. "Belief and the Will." *The Journal of Philosophy*, vol. 81 (5): 235–256. [1], [12]

—. 1989. *Laws and Symmetry*. Oxford University Press, Oxford. [2]

—. 1995. "Belief and the Problem of Ulysses and the Sirens." *Philosophical Studies*, vol. 77 (1): 7–37. [12]

Wagner, Carl. 2013. "Is Conditioning Really Incompatible with Holism?" *Journal of Philosophical Logic*, vol. 42 (2): 409–414.

Weatherson, Brian. 2013. "Disagreements, Philosophical and Otherwise." In *The Epistemology of Disagreement: New Essays*, Jennifer Lackey & David Christensen, editors. Oxford University Press, Oxford. [2]

—. ms. "Do Judgments Screen Evidence?" [2]

Weisberg, Jonathan. 2007. "Conditionalization, Reflection, and Self-Knowledge." *Philosophical Studies*, vol. 135 (2): 179–97. [13]

—. 2009. "Commutativity or Holism? A Dilemma for Conditionalizers." *British Journal for the Philosophy of Science*, vol. 60 (4): 793–812. [26]

—. 2015. "Updating, Undermining, and Independence." *The British Journal for the Philosophy of Science*, vol. 66 (1): 121–159.

White, Roger. 2006. "Problems for Dogmatism." *Philosophical Studies*, vol. 131 (3): 525–557. [10]

Williamson, Timothy. 2000. *Knowledge and its Limits*. Oxford University Press, Oxford. [4]

—. 2011. "Improbable Knowing." In *Evidentialism and its Discontents*, T. Dougherty, editor. Oxford University Press, Oxford. [6], [7]

—. 2014. "Very Improbable Knowing." *Erkenntnis*, vol. 79 (5): 971–999. [4], [6]

Zendejas Medina, Pablo. ms. "Just as Planned: Conditionalization, Externalism, and Plan Coherence." [2]