Philosophy and
Phenomenological Research

Philosophy and Phenomenological Research
Vol. XCVI No. 3, May 2018
doi: 10.1111/phpr.12329

An Accuracy Based Approach to Higher Order Evidence

MIRIAM SCHOENFIELD
The University of Texas at Austin / New York University

The aim of this paper is to apply the accuracy based approach to epistemology to the case of higher order evidence: evidence that bears on the rationality of one's beliefs. I proceed in two stages. First, I show that the accuracy based framework that is standardly used to motivate rational requirements supports *steadfastness*—a position according to which higher order evidence should have *no* impact on one's doxastic attitudes towards first order propositions. The argument for this will require a generalization of an important result by Greaves and Wallace for the claim that conditionalization maximizes expected accuracy. The generalization I provide will, among other things, allow us to apply the result to cases of self-locating evidence. In the second stage, I develop an alternative framework. Very roughly, what distinguishes the traditional approach from the alternative one is that, on the traditional picture, we're interested in evaluating the expected accuracy of *conforming* to an update procedure. On the alternative picture that I develop, instead of considering how good an update procedure is *as a plan to conform to*, we consider how good it is *as a plan to make*. I show how, given the use of strictly proper scoring rules, the alternative picture vindicates *calibrationism*: a view according to which higher order evidence should have a significant impact on our beliefs. I conclude with some thoughts about why higher order evidence poses a serious challenge for standard ways of thinking about rationality.

1. Introduction

The aim of this paper is apply accuracy based considerations to the debate about higher order evidence. To illustrate what the debate is about, it will be helpful to have a particular case in mind. So consider:

> Hypoxia: Aisha is flying her airplane on a bright Monday morning, wondering whether she has enough gasoline to fly to Hawaii. Upon looking at the dials, gauges and maps, she obtains some first order evidence E, which she knows strongly supports (say to degree 0.99) either that she has enough gas (G) or that she does not have enough gas (~G). Aisha does some complex calculations and concludes G, which is, in fact, what E supports. But she then gains some higher order evidence: she realizes that she is flying at an altitude that puts her at great risk for hypoxia, a condition that impairs one's reasoning capacities. Aisha knows that pilots who do the kind of reasoning that she just did, and who are flying at her current altitude, only reach the correct conclusion 50% of the time.[1]

How confident should Aisha be that she has enough gas? This is the sort of question that the debate about higher order evidence has centered around, and there are two answers to this question that will be the primary focus of this paper.

[1] This case is based on a case from Elga (ms).

The calibrationist thinks that Aisha's first order evidence (dials, gauges, maps) supports a 0.99 credence in G (as stipulated), but that her *total* evidence supports a 0.5 credence in G. Even though Aisha, in fact, drew the correct conclusion from her first order evidence, she can't rationally be confident that she did because of the higher order evidence she possesses concerning hypoxia. And so, say the calibrationists, her credence in G should be 0.5.[2]

I think that the calibrationist's judgment has a great deal of intuitive plausibility, but a number of philosophers have developed compelling arguments against calibrationism and have defended, instead, a position that I will call "steadfastness."[3] As applied to the case at hand:

The *steadfaster* thinks that Aisha's higher order evidence is completely irrelevant to the question of whether G, and that her total evidence supports the same doxastic attitude that her first order evidence supports: a 0.99 credence in G. After all, say the steadfasters, what possible bearing could facts about Aisha's mental states have on the likelihood that there is enough gasoline in the tank?

While the debate about higher order evidence has been unfolding, there has been an increasing interest in what is sometimes called "accuracy-first epistemology": the project of deriving rational requirements from accuracy based considerations.[4] The goal of this paper is to apply the accuracy based approach to the case of higher order evidence. In the first part of the paper, I show that the standard accuracy based approach favors steadfasting. The argument for this will require a generalization of an important result from Greaves and Wallace (2006) and a discussion of what this generalization tells us about the relation between conditionalization and self-locating evidence. In the second part of the paper, I will sketch an alternative way of thinking about epistemic rationality and argue that, unlike the standard picture, it *can* make good sense of the motivations behind calibrationism. I conclude with some thoughts about why, and how, higher order evidence challenges our standard ways of thinking about epistemic rationality.

2. Why Higher Order Evidence is Puzzling

Before getting into the nitty-gritties, let's think intuitively about which view of higher order evidence an interest in accuracy might motivate. For now, we can characterize an interest in accuracy as an interest in having high credences in truths and low credences in falsehoods.

Here is the first thing to note: Intuitively, it seems reasonable for us to expect that if Aisha successfully conforms to the steadfaster's recommendations, she will be more accurate than she would be if she successfully conformed to the calibrationist's recommendations. For we should expect that if Aisha has a 0.99 credence in G, as the steadfaster recommends, she will almost certainly have a 0.99 credence in the truth. After all, we are imagining that the configuration of the plane, as described by Aisha's evidence E,

2 Some calibrationists include Christensen (2010), Sliwa and Horowitz (2015), Vavova (forthcoming) and Elga (ms).

3 The name comes from the related debate about how we ought to respond to peer disagreement (see Christensen (2010) for a discussion of the connections between these debates). Some steadfasters include Lasonen Aarnio (2014), Titelbaum (2015) and Weatherson (ms). (Kelly's (2005) paper is also notable since it defends steadfastness with respect to peer disagreement. In the paper, Kelly is motivated by considerations that, I think, favor steadfastness concerning higher order evidence more broadly.) See also White (2009) and Schoenfield (2015a) for arguments that challenge calibrationism.

4 For an overview, see Pettigrew (2016).

makes it nearly certain that it contains enough fuel to make it to Hawaii. So if Aisha adopts a 0.5 credence, as the calibrationist recommends, she is almost certainly missing out on a terrific opportunity to have a very high credence in a truth (and, perhaps, also a terrific opportunity for a vacation in Hawaii)! Since it seems, antecedently, that we should expect steadfasters to be more accurate than calibrators, the steadfaster has a prima facie argument for the claim that steadfasting, as opposed to calibrating, is the response to the evidence that is best motivated by an interest in being accurate.

One might feel, however, as if there should be a way to resist the thought that steadfasting is what's favored by accuracy considerations, at very least, *in the right sense* of "favored by accuracy considerations." For example, it might seem that the sense in which steadfasting is favored by accuracy considerations is the same as the sense in which the rule "believe all and only the truths" (let's call this "the truth rule") is favored by accuracy considerations. Clearly, the truth rule doesn't describe the right theory of rationality, even from an accuracy-first perspective. (At very least, no accuracy-first epistemologists that I am aware of have endorsed the truth rule.) So one might think that steadfasting should be ruled out as the view motivated by accuracy-first epistemology for the very same reason that the truth rule is ruled out. But for this kind of response to be satisfying, we will want some story about *why* accuracy-first epistemologists don't endorse the truth rule as the correct theory of rationality. Otherwise, it's not clear that the steadfasting rule could be dismissed on the same grounds. So let's now consider the question: why don't accuracy-first epistemologists endorse the truth rule?

Not much is said explicitly concerning what's wrong with the truth rule in the accuracy-first literature, but implicit in this literature is the thought that we should not require agents to believe, or respond to, information that they do not possess (see, for example, Greaves and Wallace (2006, p.459)). So what accuracy-firsters want (and they *are* explicit about is this) is an *evidentialist* theory—a theory in which what an agent ought to believe is a function of her evidence alone. The problem with the truth rule, then, is that it is inconsistent with evidentialism: it requires an agent to respond to information that is not in her possession. For we can imagine plenty of cases in which an agent with evidence E lives in a world in which some proposition P is true, and cases in which an agent with evidence E lives in a world in which P is false. Evidentialism says that these agents are required to believe the same things, the truth rule says that they are required to believe different things, and so the truth rule is inconsistent with evidentialism.

Unlike the truth rule, however, steadfastness *is* consistent with evidentialism. Steadfastness simply says that if an agent has as her evidence E+H, the rational response is a 0.99 credence in G. The steadfaster is not asking Aisha to respond to information that's not in her possession. She is asking Aisha to respond to information that *is* in her possession in a particular way: by adopting a 0.99 credence in G.

Still, one might feel that there is something suspiciously truth-rule-y about the steadfaster's position. Here is one way to develop this idea: evidentialism is motivated by the thought that we should not be requiring agents to respond to information that is not in their possession. But *this* thought, one might argue, is motivated by an even more general thought: that we should not be requiring agents to do things that they're not capable of doing. And just as agents simply *can't* respond to information that is not in their possession, they also simply *can't* respond to information that is in their possession in the way that the steadfaster recommends when they are hypoxic. So the problem with steadfasting is that it's asking agents who are cognitively impaired to do something that requires cognitive capacities that they lack.

Such a proposal cannot be the right answer, at least without further elaboration. For the calibrationists themselves make a point of emphasizing that the victims of higher order evidence may be completely cognitively intact. The judgment that Aisha should reduce confidence remains even if (unbeknownst to her) she is not susceptible to hypoxia. If Aisha is not, in fact, hypoxic, though she receives evidence that makes it rational to suspect that she is, then the problem with mandating that Aisha assign a 0.99 credence in G can't be that Aisha is not able to do the calculations that would lead her to this conclusion: her cognitive capacities, in this version of the case, are in stellar condition!

If the problem with requiring Aisha to steadfast isn't that she *can't* steadfast, perhaps the thought is something along these lines: the steadfast policy, which, more generally, has an agent always adopt the attitude that her first order evidence warrants, is not a policy that we can *expect* to always be able to conform to. Indeed, the steadfasting policy tells us to have the attitudes supported by our first order evidence precisely in those cases in which our higher order evidence tells us that we will likely fail at determining what those attitudes are. Since we can't *expect*, in such circumstances, to be able to do what the steadfasting policy recommends, we can't be rationally required to conform to it.

This strikes me as an unpromising route. Consider a body of evidence that includes information to the effect that I am bad at evaluating evidence about the future outcomes of sports matches that I care about a great deal. Say that, for any given credence I adopt on such matters, 50% of the time it turns out that I am a bit overconfident due to wishful thinking, and 50% of the time I am a bit underconfident due to fear of disappointment. If my evidence includes this information, then I can't expect to successfully adopt the credences that are supported by my sports related evidence. But that doesn't mean that I am not rationally required to do so. Beliefs that are unsupported by the evidence due to wishful thinking or fear of disappointment are irrational *even if* the wishful thinkers or fearers can't help themselves *and know that they can't*. Similarly, I might not be able to expect to rationally evaluate my child's musical abilities. That doesn't mean that there is no attitude that my evidence supports about the matter. It just means that I likely will not adopt that attitude. So it can't be that the problem with responding to the evidence in the way that the steadfaster recommends is merely that we can't expect to *succeed* at conforming to her recommendations.

We are left, then, with following puzzling phenomenon: On the one hand, it seems like *especially* when being accurate is extremely important, calibrating makes sense. It's intuitive. On the other hand, it seems, we should expect steadfasters to be more accurate than calibrators. The calibrationist's claim that, given certain bodies of evidence, we are required to adopt a particular state that we antecedently expect to be less accurate than an alternative response to that evidence, is, at very least, quite out of the ordinary. I argued that if there is a problem with the steadfaster's view, it is not just that we can't always succeed at steadfasting or that we can't expect to succeed at steadfasting. Being rational can be difficult and it is not a constraint on rational requirements that we should expect to always be able to successfully follow them. So why is it so tempting to think that it's rational for Aisha to calibrate? In the second part of the paper I will address this question.

Part I—The Standard Accuracy Based Approach

The aim of Part I is to provide an argument for steadfasting using the standard accuracy based approach for deriving rational requirements. To do this, I will appeal to a result proved by Greaves and Wallace: that *conditionalization maximizes expected accuracy.* As we'll see, however, Greaves and Wallace's result does not apply universally. On some ways of conceiving of higher order evidence—namely, thinking of higher order evidence as *self-locating* evidence—cases of higher order evidence won't fall within the scope of cases to which the Greaves and Wallace result applies. So one of the things we'll have to do along the way is generalize their result so that we can apply it to self-locating propositions as well.

In brief: the reason that the standard accuracy based approach delivers the result that we should steadfast is that the standard approach evaluates an update procedure by calculating the expected accuracy of *conforming* to it. As I mentioned in the introduction, it's not intuitively surprising that we should expect conforming to the steadfaster's recommendations to lead to better results than calibrating. So why bother providing a formal proof?

There are three reasons. First, the claim that the standard accuracy driven approach favors steadfasting has important consequences. It has, for example, the consequence that the calibrationists cannot endorse the accuracy based approach to deriving rational requirements, at least as it is currently conceived, and that the accuracy-first proponents are taking on substantive commitments concerning the higher order evidence debate. Since the claim has important consequences, it is worth seeing the argument for it in detail rather than simply relying on one's intuitions. Second, as we'll see, there are some moves that the calibrationist can make to resist the thought that steadfasting is more expectedly accurate than calibrating. These moves will become apparent once we lay out the argument in detail. Finally, despite the fact that it's intuitive that steadfasting leads to more accuracy than calibrating, one might wonder *why* this is intuitive: after all, from a calibrationist's perspective, 0.5 is the credence that's supported by the total body of evidence. (They insist on this!) Why, then, *do* we expect 0.99 to be more accurate?

3. The Accuracy Framework and The Greaves and Wallace Result

In this section I will describe a few highlights of the accuracy based framework that will be relevant to both my argument for the claim that steadfasting is favored by the standard approach, and that calibrating is favored by the alternative approach that I develop at the end of the paper.

To begin, what exactly is accuracy, and how do we measure it? Intuitively, we can think of the accuracy of some credence function as its "closeness to the truth." c is maximally accurate if it assigns 1 to all truths and 0 to all falsehoods. It is minimally accurate if it assigns 1 to all falsehoods and 0 to all truths. More formally, **accuracy** is measured by a scoring rule, **A,** which takes a credence function, c, from the set of possible credence functions, C, and a possible state of the world, s, from a partition of possible states, S, and maps the credence-function/state pair to a number between 0 and 1 that represents how accurate the credence function is in that state.

$$\mathbf{A} : C \times S \rightarrow [0, 1]$$

For the purposes of this paper, I will assume (along with the others involved in this literature) that the accuracy measure we use is *strictly proper*.[5] This means that the accuracy

[5] See Greaves and Wallace (2006), Gibbard (2008), Joyce (2009), Moss (2011), Horowitz (2013), Pettigrew (2016) and Levinstein (ms). for a discussion of the motivation for using strictly proper scoring rules.

measure has the property that each probability function assigns greater *expected accuracy* to itself than to any other credence function. But what is the expected accuracy of a credence function? The expected accuracy of a credence function c relative to a probability function p is the average of the accuracy scores that c has in each state, weighted by the probability that p assigns to each such state obtaining. That is, the **expected accuracy** of credence function c relative to probability function p is:

$$\mathrm{EA}^{p}(c) = \sum_{\mathrm{s} \in \mathrm{S}} p(\mathrm{s})\mathbf{A}(c, \mathrm{s})$$

An important result for the purposes of this paper is Greaves and Wallace's (2006) argument for the claim that *conditionalizing on the (strongest) proposition one learns is the update procedure that maximizes expected accuracy*. This claim needs some unpacking. We have defined the expected accuracy of a credence function, but what is the expected accuracy of an update procedure?

Here is the basic thought: suppose you know that, at some future time t, you will learn exactly one proposition from a set of propositions $\mathcal{X}$. Let $\mathcal{X}$—the set consisting of propositions you think you might learn at t[6]—represent your future learning experience. For example, you might know that at 6pm you will see your friend and that, upon talking to her, you will learn one of the following two propositions:

Café: {Taylor's Café is open on Sundays (O), Taylor's Café is not open on Sundays (~O)}

If you know that you will learn exactly one of these propositions, we will say that the set above, labeled "*Café*, " represents your future learning experience.

Now, suppose you are interested in the proposition that the café will be open on a national holiday (H). There are lots of different ways to revise your credence in H in response to learning one of the propositions in *Café*. For example, you might decide to adopt 0.7 in H (and 0.3 in ~H) if you learn O, but 0.2 in H (and 0.8 in ~H) if you learn ~O. We'll call each such way of revising your credences in response to what you learn an "update procedure." We can represent an update procedure, U, as a function from the propositions you might learn (in this case, the members of *Café*), to probability functions. So, for example, the update procedure just described is a function U from the members of *Café* to probability functions that looks like this:

$$\begin{aligned} U(\mathrm{O}) &= p_1 \\ &\text{where } p_1(\mathrm{H}) = 0.7 \\ &\quad p_1(\sim\mathrm{H}) = 0.3 \\ U(\sim\mathrm{O}) &= p_2 \\ &\text{where } p_2(\mathrm{H}) = 0.2 \\ &\quad p_2(\sim\mathrm{H}) = 0.8 \end{aligned}$$

Greaves and Wallace say that, on their intended interpretation, you *conform* to an update procedure U if and only if you adopt $U(\mathcal{X}_{\mathrm{i}})$ (that is, the credence function that U assigns

6 More precisely, $\mathcal{X}$ is the set of propositions that you assign non-zero credence to learning at time t.

to $\mathcal{X}_i$) whenever you learn $\mathcal{X}_i$. So you conform to the update procedure U above provided that if you learn O, you adopt p_1, and if you learn ~O, you adopt p_2.

We can now examine all of the different possible update procedures in response to a learning experience like *Café*, and ask: which update procedure is such that conforming to it has greatest *expected accuracy*, where the expected accuracy of an update procedure U, relative to a probability function p, is just the weighted average of the accuracy scores that would result from conforming to the procedure.

More formally (if you're interested): Let S be the partition of states that your probability function p is defined over. Let $L(\mathcal{X})$ be the set of propositions $L(\mathcal{X}_i)$, where $L(\mathcal{X}_i)$ is the proposition that you learn $\mathcal{X}_i$ upon undergoing the learning experience in question. Since, in the cases we're imagining, you are certain that you will learn exactly one proposition from the set of propositions $\mathcal{X}$, you are certain that exactly one member of $L(\mathcal{X})$ is true. (Thus, the $L(\mathcal{X}_i)$ form a partition over S.) Now, let $U(s)$ be the credence function that you would adopt in state s if you conform to U in s. Since conforming to U involves adopting $U(\mathcal{X}_i)$ whenever you learn $\mathcal{X}_i$ (that is, whenever $L(\mathcal{X}_i)$ is true), this means that $U(s)$ will equal $U(\mathcal{X}_i)$ whenever s is a state in which you learn $\mathcal{X}_i$—that is, whenever $s \in L(\mathcal{X}_i)$. So the weighted average of the accuracy scores you would get by conforming to update procedure U relative to a probability function p is:[7]

$$\mathrm{EA}_p(U) = \sum_{s \in L(\mathcal{X})} p(s)^* \mathbf{A}(U(s),\ s)$$
$$= \sum_{L(\mathcal{X}_i) \in L(\mathcal{X})} \ \sum_{s \in L(\mathcal{X}_i)} p(s)^* \mathbf{A}(U(\mathcal{X}_i), s)$$

Now that we have the notion of the expected accuracy of an update procedure on the table, let's return to our question: which update procedure maximizes expected accuracy in response to a learning experience? Greaves and Wallace provide the following answer: *conditionalizing on the proposition you learn is the update procedure that maximizes expected accuracy*. In other words, say Greaves and Wallace, if you know that you will learn some proposition $\mathcal{X}_i$ from a set of propositions $\mathcal{X}$, the way to maximize expected accuracy is for:

$$U(\mathcal{X}_i) = p(\cdot|\mathcal{X}_i)$$

$$\text{where } p(A|B) = p(A\&B)/p(B)$$

However, it will be important for later to note that their result only shows that conditionalizing on what you learn maximizes expected accuracy when the agent satisfies the following two conditions:

> PARTITIONALITY: She is certain (prior to undergoing the learning experience) that *exactly one* of the propositions in the set of propositions that she thinks she might learn is true.

---

7 My definition of the expected accuracy of an update procedure is a slightly generalized version of the definition given by Greaves and Wallace. The connection between the two and my reason for offering the generalization is explained in Appendix 1 and will become clear once we apply the framework to higher order evidence cases.

(In other words: the set of propositions, $\mathcal{X}$, she thinks she might learn is a partition of her possibility space.)

FACTIVITY: She is certain that, for all $\mathcal{X}_i \in \mathcal{X}$, if she learns $\mathcal{X}_i$, then $\mathcal{X}_i$ is true.

(Greaves and Wallace are explicit about PARTITIONALITY. I explain why both assumptions are necessary in Appendix 1.)

4. Applying the Greaves and Wallace Result

We are now in a position to address the question: which position on higher order evidence recommends revising one's beliefs in the way that maximizes expected accuracy?

To figure out which update procedure maximizes expected accuracy in Aisha's case, we must first determine which set of propositions represents the learning experience that she will undergo on the flight. It is important to realize that, in the framework we're working with, we consider the expected accuracy of various update procedures *before* undergoing the learning experience. (I will discuss the reasons for this later.) Since Aisha flies her plane on Monday, let's imagine that, on Sunday, she is considering which update procedure maximizes expected accuracy in response to the learning experience that will take place on Monday. To do this, she considers the different propositions she might learn. What are these propositions? Let's suppose for simplicity that there are only two bodies of first order evidence (the settings of dials and gauges on the plane) that Aisha might receive: E and ~E, and that Aisha is going to either learn that her reasoning is impaired on Monday or that it's not. On one way of telling the story, then, Aisha knows, on Sunday, that on Monday she will learn exactly one of the following four propositions:

(1) E and my reasoning is impaired on Monday.

(2) E and my reasoning is not impaired on Monday.

(3) ~E and my reasoning is impaired on Monday.

(4) ~E and my reasoning is not impaired on Monday.

Let H_m be the proposition that Aisha's reasoning is impaired on Monday. We can represent the four propositions that Aisha might learn as:

$$\mathcal{F}: \{EH_m, E \sim H_m, \sim EH_m, \sim E \sim H_m\}$$

Let's stipulate that Aisha satisfies PARTITIONALITY and FACTIVITY. She knows that exactly one proposition in $\mathcal{F}$ is true, and that, whichever proposition she learns, it will be the true one. Given this, the Greaves and Wallace result tells us that, on Sunday, Aisha should regard conditionalizing on the members of $\mathcal{F}$ as the update procedure that maximizes expected accuracy.

But conditionalizing on the propositions in $\mathcal{F}$ yields *steadfastness*. The reason for this is that, *on Sunday*, Aisha will regard the quality of her reasoning capacities *on Monday* to be completely irrelevant to the question of whether there will be enough gas in the tank on the supposition that various facts about the plane obtain. So, on Sunday, Aisha's credence that G conditional on E should be the same as Aisha's credence that G conditional on E *and Aisha's reasoning is impaired on Monday*. This means that, if p_s represents Aisha's probabilities on Sunday, we'll have:

$$p_s(\mathrm{G}|\mathrm{E}) = p_s(\mathrm{G}|\mathrm{EH_m}) = 0.99$$

Thus, on the assumption that what Aisha learns in HYPOXIA is $\mathrm{EH_m}$, the accuracy optimizing update procedure is the one according to which Aisha assigns 0.99 to G in HYPOXIA, since this is the credence in G that would result from conditionalizing.

David Christensen (a calibrationist) makes a similar point in his (2010). He writes:

> So it seems that the [higher order evidence] about my being drugged produces a mismatch between my current confidence that [G] is true on the supposition that I will learn certain facts, and the confidence in [G] that I should adopt if I actually learn those facts. (p. 200)

Even if one isn't particularly interested in maximizing expected accuracy, the phenomenon described by Christensen is odd, and Christensen clearly recognizes this. Normally, how confident we should be in a proposition on the supposition that some facts obtain is the same as how confident we should be in that proposition if we go on to learn these facts. This is the central insight underlying conditionalization. But calibrationism seems to break this very natural connection: on Sunday, Aisha's credence in G on the supposition that E and she is impaired on Monday should be 0.99. But, says the calibrationist, if she actually goes on to learn these things on Monday, her credence in G should be 0.5. So after noting this puzzling phenomenon that higher order evidence gives rise to, Christensen briefly proposes an interesting way of thinking about Aisha's evidence that might avoid the problem. The suggestion is that we think of her evidence as a self-locating proposition. I turn to this proposal in the next section in which I argue that, while initially promising, it ultimately will be of no help to the calibrationist.

5. Conditionalizing on Self-Locating Evidence

On Christensen's suggestion, rather than thinking of Aisha's evidence on Monday as *E and Aisha is impaired on Monday* ($\mathrm{EH_m}$), we should think of her evidence on Monday as the self-locating proposition:[8] *E and I'm impaired now* ($\mathrm{EH_{now}}$). A useful heuristic for distinguishing self-locating from non-self-locating propositions is to note that $\mathrm{EH_m}$ is the sort of proposition one might learn by reading "the encyclopedia of the world": a book that describes, in third personal terms, everything that happens in the past, present and future. For example, it might say: "And then, on Monday June 24th 2015 at 3:31 pm, Aisha flew a plane and was hypoxic." The propositions in the encyclopedia are not only true now—they're true *forever*. No matter what happens in the future, it will always and forever be the case that on Monday June 24th... The encyclopedia of the world, however, has one important limitation: it doesn't tell you who *you* are in the story, or what time it *now* is. It doesn't come with a "you are here" sticker. One reason it doesn't contain these important bits of information is that it's meant to be an encyclopedia that everyone can read and learn from forever. Since what time it is *now* is constantly

[8] Such propositions are sometimes described as "centered," "*de se*," or "essentially indexical."

changing, and the encyclopedia (it's an old fashioned one) can't keep rewriting itself, it simply cannot contain this sort of information. For our purposes, we can think of self-locating propositions as those that you can't learn from the encyclopedia of the world, and we'll be focusing specifically on temporal propositions of this sort. What's crucial about these propositions is that they change their truth value over time.

Why might thinking of Aisha's evidence as consisting of such propositions help the calibrationist? The thought is that, just as the practical import of "S is trailing sugar from his cart at time t" (that's unfortunate, but not my problem) is different from the practical import of "*I* am trailing sugar from my cart *now*" (better do something about this!),[9] the *epistemic* import of "S's reasoning is impaired at time t" (aww…poor S) is different from the epistemic import of "*my* reasoning is impaired *now*" (better revise my credences!).

To see how this suggestion can be developed, let "H_{now}" be the proposition that my reasoning is impaired *now*. (Formally, we can represent H_{now} as the set of centered worlds in which the center's reasoning is impaired, but this formal characterization is not necessary for what follows.) The suggestion hinted at by Christensen is that, while perhaps conditionalizing on H_m doesn't yield calibrationist results, conditionalizing on H_{now} does. It's true, the calibrationist can grant, that Aisha, on Sunday, shouldn't regard her reasoning capacities on *Monday* to be relevant to how much gas is in her tank supposing that E is the case. But even on Sunday, says the calibrationist, Aisha should consider her *current* reasoning capacities to be relevant to her degree of confidence in G given E. Thus, the calibrationist may claim that Aisha's conditional probabilities on Sunday should look like this:

- $p_s(G|E) = \text{high}$
- $p_s(G|EH_{now}) = 0.5$.

If Aisha's Sunday conditional probabilities should be as described above (and I'll just grant the calibrationist that this is the case), then the calibrationist can claim that when, on Monday, Aisha learns EH_{now}, conditionalization will recommend that she move her credence to 0.5.

Recall that one reason that we're interested in whether we can model the calibrationist as a conditionalizer is that we already know that, in a broad range of cases (those in which the agent satisfies PARTITIONALITY and FACTIVITY), conditionalizing on our evidence maximizes expected accuracy. Now, suppose that the calibrationist can make the case that the four propositions Aisha might learn on Monday are:

$$\mathcal{F}^* : \{EH_{now}, E \sim H_{now}, \sim EH_{now}, \sim E \sim H_{now}\}$$

If Aisha's learning experience on Monday should be represented by $\mathcal{F}^*$, then, assuming that PARTITIONALITY and FACTIVITY are satisfied, she should, on Sunday, regard conditionalizing on the members of $\mathcal{F}^*$ as the update procedure that maximizes expected accuracy. This might give the calibrationist everything she needs to defend her approach to higher order evidence. It *is*, she might claim, the result of conditionalizing, and so it is, in fact,

[9] This example comes from Perry (1979).

the approach that maximizes expected accuracy. One simply has to be clear about the fact that the relevant evidence in higher order evidence cases is self-locating.

There are a variety of worries one might have with this strategy. The first is that, as many authors have pointed out, conditionalizing is clearly not always the right (or accuracy-optimizing) way to revise one's credences when we consider self-locating propositions.[10] To see this, suppose that I'm in a waiting room, entertaining myself by staring at the clock. I assign credence 1 to the proposition *it's now 2:30*. If, in my attempt to be a good Bayesian, when I see that the clock reads 2:31, I conditionalize on the proposition *it's now 2:31*, I will end up with the absurd (and false) belief that it's now 2:30 *and* it's now 2:31. (This is because, in general, if one starts out with credence 1 in P, conditionalizing on Q results in credence 1 in P&Q.) If conditionalizing leads to absurd results in cases of self-locating evidence, then the proposal that we think of calibrating as the result of conditionalizing on self-locating evidence won't look very attractive.

However, dismissing the "self-locating evidence" solution because self-locating evidence sometimes causes trouble for conditionalization would be overly hasty. This is because it's not clear that conditionalizing is *always* the wrong way to revise one's credences in light of self-locating evidence. To see this, consider the following variation of the Sleeping Beauty case[11] (this is a version of Darren Bradley's (2011) Red-and-Blue light case):

> Fairies and Demons: Sleeping Beauty is going to be awoken and put back to sleep ten times, beginning on Monday morning. After each awakening, she will encounter either a friendly fairy or an evil demon and then her memory of the awakening will be erased before she is put back to sleep. How many fairies or demons she will see will be determined by whether she was blessed or cursed at the time of her birth. If she was blessed, she will see a friendly fairy on nine days and an evil demon on one day, but if she was cursed, she will see an evil demon on nine days and a friendly fairy on one day. Beauty knows all of this before going to sleep and her initial credence that she was blessed at the time of her birth is 0.5.

Now, suppose that Beauty wakes up and sees a fairy. How confident should she be at this point that she was blessed? Bradley thinks that seeing the fairy should increase Beauty's confidence in the Blessed hypothesis from 0.5 to 0.9. And, he claims, conditionalizing on the self-locating proposition *I see a friendly fairy today* is exactly what is necessary to deliver the desired result. Here's why: It's plausible that, upon waking, but *before seeing either the fairy or demon*, the following should be true of Beauty:

$$\Pr(\text{I see a fairy today}|\text{Blessed}) = 0.9$$

$$\Pr(\text{I see a fairy today}|\text{Cursed}) = 0.1$$

If Beauty starts out with a prior probability of 0.5 in both Blessed and Cursed, it follows from Bayes theorem[12] that:

---

10 See, for example, Arntzenius (2003), Hitchcock (2004), Halpern (2005), Meacham (2008), Titelbaum (2008), Bradley (2011) and Moss (2012).

11 The original Sleeping Beauty case was introduced to the philosophical literature by Elga (2000).

12 Let "I see a fairy today" be "F." Then:
Pr(Blessed|F) = Pr(F|Blessed)Pr(Blessed)/ [Pr(F|Blessed)Pr(Blessed) +Pr(F|~Blessed)Pr(~Blessed)] = (0.9)(0.5) / [(0.9)(0.5) + (0.1)(0.5)] = 0.9

$$\Pr(\text{Blessed}|\text{I see a fairy today}) = 0.9$$

And this is why, says Bradley, once she sees the fairy, she should have a 0.9 credence in Blessed.

Bradley points out that, in order to get the result that Beauty should assign a 0.9 credence to being blessed upon seeing the fairy by conditionalization, it's important that the proposition being conditionalized on is a self-locating one—not one that can be learned from reading the encyclopedia of the world. For if we considered instead propositions like: *Beauty sees a fairy on one of the awakenings*, conditionalizing won't yield the desired result. Since Beauty is certain that, whether she is blessed or cursed, she will see a friendly fairy on *one* of the awakenings, that she sees a friendly fairy on one of the awakenings provides no evidence either for being blessed or for being cursed. Thus, to get the desired result in cases like Fairies and Demons, says Bradley, we need to conditionalize on a self-locating proposition like: "I see a friendly fairy *today*."[13]

Let's put aside the friendly fairies for now and return to hypoxic pilots. What the example above illustrates is that conditionalizing on self-locating propositions seems, to at least some people, to be the appropriate response to one's evidence at least some of the time. If conditionalizing on self-locating evidence is sometimes appropriate, then higher order evidence cases may be of this sort. Thus, the calibrationist may be able to argue that her response involves conditionalizing on self-locating evidence *appropriately*.

Now, of course, a full-blooded version of this response would require some way of distinguishing the cases in which conditionalizing on self-locating evidence is appropriate and cases in which it isn't, and then arguing that higher order evidence cases are cases in which conditionalizing is warranted. The question of how, in general, to revise in light of self-locating evidence is incredibly controversial[14] (and I'll offer my own proposal near the end of this section). So, rather than survey the vast literature on this subject, I will simply focus on Bradley's account since his account is the one that will be most helpful to the calibrationist. Indeed, Bradley offers a theory for when conditionalizing on self-locating evidence is appropriate that seems to favor exactly what the calibrationist is suggesting.

Bradley argues that we should conditionalize on self-locating evidence when, and only when, what we learn is a *discovery*. Very roughly, this is because, Bradley claims, conditionalization involves "eliminating false possibilities and zooming in on the truth" (p. 394). When you're staring at the clock ticking in the waiting room, there are no surprises to be had. When you see the clock change from 2:30 to 2:31, you don't say to yourself: "Well my goodness! I thought it was 2:30—what a surprise!" Since there are no surprises, says Bradley, there's no "eliminating false possibilities" (in the relevant sense) and so conditionalization shouldn't be deployed. Thus, in the cases that he calls cases of *belief mutation*—belief change in virtue of a change in the *truth value* of the content of the belief (like the clock case)—one shouldn't conditionalize. However, in

13 I use this example, rather than the original sleeping beauty case, because the judgment in the original case that requires conditionalizing on self-locating propositions (the "thirder" judgment) is highly controversial. The judgment that Beauty should assign a 0.9 credence in this case, however, will be accepted by both thirders and at least some halfers (such as Bradley himself).

14 See note 10 for references.

cases of *discovery*—belief change in virtue of the discovery of the truth of the content of the belief (e.g. I wonder whether I'll see a fairy or a demon on this awakening—ah-ha! It's a fairy!)—Bradley thinks that one ought to conditionalize on the self-locating evidence.

What about Aisha's case? Aisha's case can certainly be thought of as a case of belief discovery in Bradley's sense. She doesn't know on Sunday whether she will be impaired on the flight, and she then discovers "Ah-ha! I'm hypoxic!" If we can think of calibrating as the result of conditionalizing on this newly *discovered* self-locating evidence, and Bradley is right that, in any case of belief "discovery," conditionalization is appropriate, then the calibrationist may be able to give a fuller story about why calibrating makes sense: we get self-locating evidence, conditionalizing on it results in the calibrationist's recommendations, and conditionalizing on this evidence is the right policy because the new self-locating information is *discovered* and not merely the result of *mutation*.

This strategy for motivating calibrationism, however, will not succeed, at least if we're interested in finding the response to higher order evidence that maximizes expected accuracy. This is because, if we're motivated by accuracy considerations, then whether or not we should conditionalize on self-locating evidence is going to ultimately depend on whether doing so maximizes expected accuracy. I will now argue that conditionalizing on self-locating evidence *doesn't* in general maximize expected accuracy in cases of belief discovery. So Bradley's way of determining when one should conditionalize doesn't correspond to what would be warranted by an interest in accuracy.[15] I will show that the update procedure that *does*, in general, maximize expected accuracy in cases of self-locating evidence is not conditionalization, but what I will call *conditionalization**. Conditionalization*, it turns out, yields steadfastness, even if we think of Aisha as gaining self-locating evidence. Thus, even if the calibrationist's recommendation can be thought of as a kind of conditionalizing, it's the wrong kind of conditionalizing, at least on the standard model connecting rationality and accuracy, since it's not the kind of conditionalizing that maximizes expected accuracy.

Here's the reason that conditionalizing on self-locating evidence doesn't, in general, maximize expected accuracy in cases of discovery: Recall that, as always, the expected accuracy of an update procedure is calculated *prior* to getting the evidence (in our case, this is Sunday).[16] Recall also that the Greaves and Wallace result only applies when the agent satisfies PARTITIONALITY and FACTIVITY. The problem is that, since self-locating temporal propositions don't come with truth values attached to them—they are only true *relative to a particular time*—FACTIVITY as applied to self-locating propositions is underpsecified.[17] FACTIVITY says that an agent, prior to undergoing the learning experience, must be certain that if she learns P, then P is true. But if P is self-locating (temporally), we must ask, true *when*? In fact, for the Greaves and Wallace result to apply in cases of self-locating propositions what's needed is:

---

15 It's also worth noting that we may be able to modify Aisha's case so that it looks more like a mutation case than a discovery case. But I will not pursue this line here.

16 It's worth noting that no calculating has to *actually* take place on Sunday. We can think about what update procedure has greatest expected accuracy relative to the set of credences that would be rational for Aisha to have on Sunday, even if Aisha never considers, on Sunday, the question of how likely G is to be true given E. (Credences, in general, are normally thought of dispositionally and don't require consciously entertaining the proposition in question.)

17 Thanks to Cameron Domenico Kirk-Giannini for helping me clarify this point.

FACTIVITY$_{\text{indexed}}$: The agent is certain at t_0 that, if she learns P upon undergoing her future learning experience, then P is true *at* t_0.

where t_0 is the time at which the agent is calculating the expected accuracy of her update procedure. That this is what's needed can be seen clearly in the argument given in Appendix 1 for why the Greaves and Wallace result relies on FACTIVITY, but we can see the intuitive idea by returning to our case of Taylor's café. Suppose, for example, that it's a Sunday and I think: supposing that Taylor's is open *today* (T), it's reasonably likely that it's open on national holidays (H). I *don't* think, however, that supposing that I learn *tomorrow* (Monday) the self-locating proposition *Taylor's is open today*, then it's reasonably likely that it's open on national holidays. For if I think that I'm going to having a learning experience tomorrow in which I might learn T, then I'll now think that I'll learn T tomorrow if and only if Taylor's is open on *Monday*, and that Taylor's is open on Monday, we may suppose, is completely irrelevant to the question of whether it is open on national holidays. Thus, in general, if P is self-locating, I'll only want to conditionalize on P in the future if I think that if I'll learn P *in the future*, then P is true *now*.

But Aisha doesn't think on Sunday that if she learns on Monday that she's impaired, then she's impaired *now* (on Sunday). So if we think of Aisha's learning experience on Monday as involving self-locating propositions like "I am impaired now," then Aisha doesn't satisfy FACTIVITY$_{\text{indexed}}$ on Sunday, and the Greaves and Wallace result doesn't apply. This is true despite the fact that what she learns is a "discovery" in Bradley's sense.

So what does, in general, maximize expected accuracy in cases in which we learn self-locating evidence (be they discoveries or mutations)? To figure this out we need a more general result than the one that Greaves and Wallace provide: we need to know which update procedures maximize expected accuracy when the conditions necessary for their result (such as FACTIVITY$_{\text{indexed}}$) are not satisfied. Here is the answer:

Generalized CondMax: Suppose you are certain that you are going to learn exactly one proposition from a set of propositions $\mathcal{X}$ at time t. Let $L(\mathcal{X}_i)$ be the proposition that $\mathcal{X}_i$ is the proposition learned upon undergoing the learning experience at t. The update procedure that maximizes expected accuracy in response to $\mathcal{X}$, relative to probability function p, is the update procedure that assigns, to each $\mathcal{X}_i$, $p(\cdot|L(\mathcal{X}_i))$.

I will call the update procedure that has us adopt $p(\cdot|L(\mathcal{X}_i))$, upon learning $\mathcal{X}_i$, "conditionalization*." The proof of the result in Appendix 2.[18]

What Generalized CondMax says is the following: in any case in which you know that you will, at some future time, undergo a learning experience during which you will learn exactly one proposition from a set of propositions $\mathcal{X}$, the following procedure is the one that maximizes expected accuracy: upon learning $\mathcal{X}_i$, conditionalize on the proposition *that* $\mathcal{X}_i$ *was the proposition learned upon undergoing the learning experience*. The reason that this is a *generalization* of the Greaves and Wallace result is that the cases in which PARTITIONALITY and FACTIVITY$_{\text{indexed}}$ are satisfied turn out to be exactly those cases in which the agent antecedently regards $\mathcal{X}_i$ and $L(\mathcal{X}_i)$ as equivalent

[18] I elaborate more on this result and its implications in Schoenfield (forthcoming*a*).

(for a proof, see Appendix 1). And so, when PARTITIONALITY and FACTIVITY$_{\text{indexed}}$ are satisfied, conditionalizing on $\mathcal{X}_i$ (ordinary conditionalization) and conditionalizing on L($\mathcal{X}_i$) (what I'm calling "conditionalization*") amount to the very same thing.

Let's apply this result to Aisha's case. We already saw that if we think of Aisha as learning self-locating propositions on Monday like *E and I am impaired now*, then FACTIVITY$_{\text{indexed}}$ isn't satisfied. Thus, the update procedure that maximizes expected accuracy isn't one that has Aisha conditionalize on EH_{now}, but rather one that has her conditionalize on *L(EH_{now})*. So now the question becomes: what is the result of conditionalizing on L(EH_{now})?

The first thing to note is that, even if some learned proposition $\mathcal{X}_i$ is a self-locating proposition—one that you can't learn about from the encyclopedia of the world—L($\mathcal{X}_i$) need not itself be self-locating. It very well may be found in the encyclopedia of the world. In Aisha's case, L(EH_{now}) is, indeed, a non-self-locating proposition. It is the following proposition: Upon undergoing the learning experience on Monday, Aisha learns a proposition that is true on Monday if and only if E is true and her reasoning capacities are impaired on Monday. And Generalized CondMax tells us that conditionalizing on this *non-self-locating* proposition is what maximizes expected accuracy. We can now see why conditionalizing on L(EH_{now}) yields steadfasting. For (assuming Aisha knows that she will only gain accurate information) L(EH_{now}) is going to be true if and only if E is true and she is impaired on Monday: that is, if and only if EH_m is true. We've already seen that conditionalizing on EH_m yields steadfastness. It follows that conditionalizing on L(EH_{now}) also yields steadfastness.

In sum, moving to self-locating propositions doesn't help the calibrationist. Even if the calibrationist thinks that Aisha receives self-locating evidence on Monday, the update procedure that maximizes expected accuracy *in response to that self-locating evidence* is not the one that has her conditionalize on the self-locating propositions she learns. It is, rather, one that has her conditionalize on the *non-self-locating* proposition *that she learns* such-and-such self-locating proposition upon undergoing her learning experience on Monday. And this, as we saw, yields steadfastness.

More generally, the answer to the question: when should one conditionalize on self-locating evidence isn't Bradley's answer (that one should conditionalize on the learned proposition whenever the proposition constitutes a genuine discovery). The answer is rather: whenever PARTITIONALITY and FACTIVITY $_{\text{indexed}}$ are satisfied. As I mentioned, PARTITIONALITY and FACTIVITY$_{\text{indexed}}$ are satisfied in exactly those cases in which the agent is antecedently certain that, for each proposition $\mathcal{X}_i$ that she might learn, $\mathcal{X}_i \leftrightarrow$ L($\mathcal{X}_i$) (see Appendix 1 for a proof). Thus, one should only conditionalize on a self-locating proposition, $\mathcal{X}_i$, when one antecedently regards $\mathcal{X}_i$ as equivalent to L($\mathcal{X}_i$). In all other cases, the proper response to learning the self-locating proposition $\mathcal{X}_i$, is to conditionalize on L($\mathcal{X}_i$).

Note that the generalized result doesn't entail that we should *never* conditionalize on self-locating propositions. For the generalized result yields precisely Bradley's desired verdict in the FAIRIES AND DEMONS case: that Beauty should conditionalize on *I see a fairy today* and so assign a 0.9 credence to being blessed upon encountering a fairy. To see this, let's consider Beauty's informational state immediately prior to learning whether there's a fairy or a demon. This will be in the morning, immediately upon awakening. At this time (assuming Beauty is certain that she won't gain false information) Beauty will be certain that:

If (in a moment) I learn *there's a fairy today*, then there's a fairy today.

In other words, $\text{Factivity}_{\text{indexed}}$ is satisfied. This is because, despite the fact that "there's a fairy today" is self-locating, both the time before the learning, and the time after the learning, have the same index ("today"). For this reason, conditionalizing on *there's a fairy today* does, indeed, maximize expected accuracy. Thus, the claim that one should conditionalize on self-locating evidence when and only when Partitionality and $\text{Factivity}_{\text{indexed}}$ are satisfied, (and that, otherwise, one should conditionalize*), is both theoretically motivated and yields intuitively plausible results.[19]

At this point, the calibrationist may respond as follows: "All you've shown me is that, on *Sunday*, I should think that the best update procedure *for Monday* is the steadfast one. But you haven't shown me that, *once I've received my evidence on Monday*, I should think that the best update procedure is the steadfast one. For once Monday rolls around, and I've received my evidence, I *will* satisfy $\text{Factivity}_{\text{indexed}}$: I will think that I've learned EH_{now} just in case I'm impaired *now*.[20] This means that, on Monday, it will be true from my perspective that conditionalizing on my evidence *does* maximize expected accuracy. And don't you think that, on Monday, I should be more concerned with what my Monday credences tell me maximizes expected accuracy, than what my Sunday ones recommend?"

There's a sense in which the calibrationist is exactly right. Call Aisha's probablility function on Monday "p_m." And say that her total evidence on Monday is M. Assuming she assigns 1 to M, the credence function that Aisha should regard as maximizing expected accuracy on Monday is indeed, $p_m(\cdot|M)$. But that's just because $p_m(\cdot|M)$ equals p_m (a result of the fact that Aisha assigns 1 to M) and, since we're using strictly proper scoring rules, every probability function maximizes expected accuracy relative to itself. Thus, since p_m maximizes expected accuracy relative to p_m, $p_m(\cdot|M)$ also maximizes expected accuracy relative to p_m.

What this demonstrates is that accuracy based considerations don't deliver any particularly interesting results about how to respond to higher order evidence if we think of the question as follows: what credence function maximizes expected accuracy *relative to the credence function that you have adopted* upon receiving the higher order evidence? No matter which update procedure you use, *once you've updated*, your new probability function will regard itself as maximizing expected accuracy.

---

19 What does this theory tell us about the ordinary Sleeping Beauty case? The framework so far yields no straightforward answer. In Fairies and Demons there is a time (Monday morning—immediately upon waking up) in which Beauty knows that she will undergo a specific learning experience (learning that there's a fairy or demon), and so she can calculate the expected accuracy of various update procedures with respect to that specific learning experience, relative to her Monday morning probability function. In the ordinary Sleeping Beauty case, however, we're interested in how to respond to the *awakening itself*. So the only time before the learning experience which we can fixate on is Sunday evening. But now we must figure out what question we're asking on Sunday. Are we aiming to maximize the expected accuracy of Beauty's credence on her *first* awakening? Her *total* expected accuracy? Her *average* expected accuracy? In cases in which multiple learning experiences are involved, what one should do, from an accuracy perspective, will frequently depend on exactly which quantity one is aiming to maximize, and, in the traditional Sleeping Beauty problem, average expected accuracy and total expected accuracy do indeed come apart (see Kierland and Monton (2005)).

20 Let's assume for simplicity that, on Sunday, Aisha knows that she will be on the plane (and potentially impaired) at exactly 12am on Monday and that at 12am she will receive the body of evidence telling her both whether E and whether she is impaired. This is just to assure that there's no time on Monday before Aisha has received the relevant evidence.

The project of thinking about which update procedures maximize expected accuracy is, essentially, a diachronic one. At very least, it is one that spans more than one probability function. The question is: relative to p_1, which probability function should I hope to adopt if I learn E? Suppose that the answer is p_2. Once you've learned E, and you've adopted some other probability function, say p_3, it is always open to you to say: "p_3 maximizes expected accuracy relative to itself! It's true that, from the perspective of p_1, p_2, rather than p_3, is the function I would have hoped to adopt if I learned E. But why should I care how things look from the perspective of p_1 if my current perspective is that of p_3?" I will not attempt to respond to this challenge. All I will point out is that the thought that underlies the standard accuracy based approach for deriving rational requirements, like the argument for the claim that it's rational to conditionalize, is that it's rational to have the credences that some *prior* probability function would have regarded as maximizing expected accuracy. If this is the picture of how rational requirements and expected accuracy considerations relate to one another, steadfastness wins. In the next section of the paper I will argue that there is an alternative way of deriving rational requirements from accuracy based considerations that does, indeed, favor calibrationism.

Part II—An Alternative Approach

6. The Planning Framework

Let's set aside rationality for a moment and think about the following activity: *deliberating about what to believe*. Deliberating about what to believe is something we do all the time. We may think about what to believe in our current situation, but we also sometimes consider what to believe in possible or future situations. A scientist, for example, may wonder what to conclude about theory T if her experiment delivers result R. I might consider how confident to be in my diagnosis of what's wrong with my car if I learn that my friend disagrees with me. Sometimes, the outcome of this deliberation is that we settle for ourselves the question of what to believe in the circumstance in question. Perhaps, for example, I settle on suspending judgment if I learn that my friend disagrees with me about the car. When we settle our deliberations about what to believe in a certain way, I will say that we have made a *doxastic plan*.[21] We can think of update procedures (functions from what we might learn to belief states) as representing doxastic plans.

On the standard accuracy based picture, we evaluate an update procedure by evaluating the expected accuracy of the credences that result from *conforming* to that procedure. But if we're deliberating about which doxastic plan to adopt, we might, instead, be interested in how accurate we expect to be as a result of *planning* to update in a certain way. Why is this? Let me begin with an illustration from practical planning.

Suppose that I am planning my vacation and I am considering two possibilities: spending my vacation camping in the woods or spending my vacation on the moon.

[21] The claim that we engage in doxastic planning doesn't presuppose voluntarism about belief. See Schafer (2014) (who is largely responsible for introducing the connection between epistemic rationality and doxastic planning into recent literature) for discussion of this issue. For further discussion of doxastic planning and its implications see Greco and Hedden (forthcoming), Schoenfield (2015b) and Schoenfield (forthcoming*b*). (See also Greco (2015) and Steele (forthcoming) who don't explicitly use the language of doxastic planning, but are naturally interpreted as appealing to similar considerations.)

Clearly, vacationing on the moon would be more exciting than camping. Nonetheless, the moon plan is worse than the camping plan. Why is this?

One might claim that I simply can't plan to go to the moon because I don't believe that I will, or can, conform to the moon plan. But such an explanation would appeal to controversial principles about planning, such as the principle that says that in order to plan to ϕ you must believe that you can, or will, ϕ. These are principles that I would like to remain neutral about. Therefore, I prefer to appeal to the following very minimal thing that we can say about the moon plan, which suffices to explain its badness: even supposing that I *can* make the plan, I can't expect anything good to come of it. (In fact, I can probably expect something bad to come of it, like spending my vacation moping around at home, feeling defeated by my failure.) On the other hand, if I plan to go camping, the likely result is that I go camping and have a terrific time. So while I can expect that the result of *conforming* to the moon plan will be better than the result of conforming to the camping plan, I can also expect that the result of *making* the moon plan (again, assuming I *can* make such a plan) will be worse than the result of making the camping plan.

We can apply the distinction between evaluating the results of *conforming* to a plan and evaluating the results of *making* a plan to doxastic planning. The crucial difference between these two activities is that, in the latter case, we can take into account the possibility that we'll fail, and how bad such failures will be. With this in mind, I will argue that *planning* to calibrate does as well, or better, expected accuracy wise, than planning to update in any of the ways that have been proposed in the literature.[22]

First, let's define the expected accuracy of *planning* to *update in accord with U*. Let T be a partition over the possibilities in which the agent expects to revise her belief state using plan U. And let T be sufficiently fine grained so as to determine for each t∈T: (a) what credence function the agent adopts in t and (b) how accurate that credence function is. Let "PL^U" represent the plan to update in accord with U and let "$\mathrm{PL}^U(\mathrm{t})$" be the credence function that the agent adopts in t (which, recall, is one of the worlds in which she makes the plan to update in accord with U). We can define the expected accuracy of planning to U, relative to a probability distribution p over T as the weighted average of the accuracy scores that the agent receives in each world in which she makes plan U. That is:

$$\mathrm{EA}^p(\mathrm{PL}^U) = \sum_{\mathrm{t}\in\mathrm{T}} p(\mathrm{t})\mathbf{A}(\mathrm{PL}^U(\mathrm{t}), \mathrm{t})$$

We can now ask: what is the expected accuracy of planning to steadfast?

If you plan to steadfast, then you plan to adopt the credences supported by your first order evidence, even if you have higher order evidence suggesting that you're impaired. In the hypoxia case, this means that you plan to assign a 0.99 credence to the proposition that the first order evidence best supports (and, let's suppose, a 0.01 credence to its negation). Once again, you might think that one simply *can't* make such a plan because one can't rationally believe that one will conform to it. But, as I mentioned earlier, I do not want to rely on any controversial principles about planning, nor do I need to.

Instead, let's assume for the sake of argument that one *can* plan to steadfast, and focus on the expected results of making such a plan. The important thing to realize is that,

22 For an approach that is similar in spirit, but applied to the debate about peer disagreement, see Steele (forthcoming).

even if you *plan* to steadfast in cases in which you're hypoxic, you should expect that the judgments you will actually make on the basis of the first order evidence will be correct only 50% of the time.[23] Thus, you should expect that, if you plan to steadfast in cases in which you're hypoxic, 50% of the time you'll assign a 0.99 credence to the truth (and .01 to the falsehood) and get quite a high accuracy score, and 50% of the time you'll assign a 0.99 credence to the falsehood (and .01 to the truth) and get quite a low accuracy score.[24]

To be more precise about this, let's restrict our attention to your credences in G and ~G. And, for any r ∈ [0,1], let "*r*" (italicized) be the probability function that assigns r to G and 1-r to ~G. Thus, *.99* is the function that assigns a 0.99 credence to G and a 0.01 credence to ~G.

Assuming that our accuracy measure assigns the same value to having credence r in the truth and 1-r in the falsehood, whether the truth is G or ~G, we can represent the accuracy score that one would get for having credence r in the truth about whether G and credence 1-r in the falsehood as **A**(*r*, G). Similarly, assuming that the accuracy measure assigns the same value to having credence r in the falsehood and 1-r in the truth, whether the truth is G or ~G, we can represent the accuracy score that one would get for having credence r in the falsehood and 1-r in the truth as **A**(*r*, ~G). Thus:

Score for r in the truth and 1-r in the falsehood:	**A**(*r*, G)
Score for r in the falsehood and 1-r in the truth:	**A**(*r*, ∼ G)

Since, as noted above, if you plan to steadfast you should expect that 50% of the time you will assign 0.99 to the truth and 50% of the time you will assign 0.99 to the falsehood, the expected accuracy of planning to steadfast is:

$$\text{(1)} \qquad \mathrm{EA}(\mathrm{PL}^{\mathrm{stead}}) = (0.5)\mathbf{A}(.99, \mathrm{G}) + (0.5)\mathbf{A}(.99, \sim\mathrm{G}).$$

Since we are supposing that the relevant impairment won't affect your ability to do what calibrationism recommends, we can expect that the result of planning to calibrate is that you *will* calibrate. This will involve adopting the credence function *.5* both when G

[23] The proposition that you judge is the proposition from the relevant partition (in our case: {G,~G}) that you were (or would be) most confident in on the basis of the first order evidence alone. I borrow this term from Sliwa and Horowitz (2015) and Weatherson (ms).

[24] I am assuming here that the result of planning to steadfast will involve assigning a 0.99 credence to the proposition that you judge (see previous note) on the basis of the first order evidence. You might question whether this is actually what an agent who planned to steadfast would do. Perhaps, rather than using her judgments, she would think: "Oh my! My plan was to assign a 0.99 credence to the proposition that the first order evidence supports, but, because I'm likely hypoxic, I don't know what proposition this is! I better just stick to my initial 0.5 credence." If this is what one expects the result of planning to steadfast will be, then the expected result of planning to steadfast will be exactly the same as the expected result of planning to calibrate. On this picture, the plan to steadfast would still have no *advantage* over the plan to calibrate and it would also, I think, be somewhat misleading to offer steadfasting as an *alternative* to calibrating in the plan-making context.

is true and when G is false.[25] Thus, the expected accuracy of planning to calibrate is:

$$\text{(2)} \qquad \mathrm{EA}(\mathrm{PL}^{\mathrm{cal}}) = \mathbf{A}(.5, \mathrm{G}) = \mathbf{A}(.5, \sim\mathrm{G}).$$

We can rewrite (2) as:

$$\text{(3)} \qquad \mathrm{EA}(\mathrm{PL}^{\mathrm{cal}}) = (0.5)\mathbf{A}(.5, \mathrm{G}) + (0.5)\mathbf{A}(.5, \sim\mathrm{G})$$

We can think of planning to steadfast as being epistemically *risky*: making the plan gives you a 50% chance at a terrific accuracy score and a 50% chance at a dreadful one. Planning to calibrate, on the other hand, can be seen as epistemically *conservative*. You get a guaranteed middling level accuracy score. So should we be risky or conservative? Answer: We should be conservative, and the argument for this will appeal to the fact that we're using a strictly proper accuracy measure.[26] To see why any such scoring rule will favor the conservative plan, it will be helpful to revisit the implications of an accuracy measure being strictly proper.

Consider Claire whose credence function is *c* (and recall that this means that Claire assigns credence c to G and credence 1-c to ~G and that we are restricting our attention to Claire's credences in G and ~G). Recall that, according to the definition of expected accuracy, the expected accuracy that *c* assigns to a credence function *x* is:

$$\text{(4)} \qquad \mathrm{EA}^{c}(x) = (\mathrm{c})\mathbf{A}(x, \mathrm{G}) + (1 - \mathrm{c})\mathbf{A}(x, \sim\mathrm{G})$$

If **A** is strictly proper, it follows that (4) is maximized when *x* assigns credence c to G and credence 1-c to ~G. That is, Claire thinks that assigning her *own* credences to G and to ~G will maximize expected accuracy.

Since:

$$\text{(5)} \qquad (0.5)\mathbf{A}(x, \mathrm{G}) + (0.5)\mathbf{A}(x, \sim\mathrm{G})$$

is just an instance of (4), it follows that (5) is maximized when *x* is a function that assigns a 0.5 credence to each of G and ~G—that is, when *x* = *.5*.

Why is this relevant to the expected accuracy of the calibrationist's and steadfaster's plans? Note that both (1) and (3) are instances of (5). The difference between them is that (1) plugs in *.99* for *x* while (3) plugs in *.5* for *x*. Since (5) is maximized when *x* = *.5*, (3) must be greater than (1). Thus, the expected accuracy of *planning* to calibrate is greater than the expected accuracy of *planning* to be steadfast.

The calibrationist plan will also do better than planning to assign any other credence to the proposition best supported by the first order evidence in cases in which the judgments you form on the basis of the first order evidence are only correct 50% of the time. For example, suppose you planned to respond to the hypoxia case by taking a compromise position. Rather than assigning 0.99 to the proposition that the first order evidence supports, you'll assign, say, a 0.7 credence to the proposition that the first order evidence

25 We could also, if we wanted to, complicate the picture by imagining hypoxic impairments that *do* impact one's ability to calibrate, but we needn't delve into these complications for the purposes of the case at hand, so I will set them aside.

26 I am simply assuming here that the measure is strictly proper and not arguing for this. See note 5.

best supports and a 0.3 credence to its negation. (This kind of position has been called by Kelly (2010) "the total evidence" view.) Calibrationism will beat any such plan. For you should expect that the result of making this plan will be that 50% of the time you'll have a 0.7 credence in the truth and 50% of the time you'll have a 0.7 credence in the falsehood. More generally, the expected accuracy of making a "total evidence" plan will be:

$$(0.5)\mathbf{A}(k, \mathrm{G}) + (0.5)\mathbf{A}(k, \sim\mathrm{G}) \tag{6}$$

where k is the credence that you plan to assign to the proposition that the first order evidence best supports. But, once again, since we're using a strictly proper accuracy measure, $k=.5$ is the value that maximizes (6).

The calibrationist position can also be generalized to cases in which you expect that, due to some cognitive impairment, you will be *somewhat* unreliable, but that you will still do better than chance. Imagine learning that the chance of a hypoxic pilot arriving at a correct judgment on the basis of her first order evidence is r. What credence should the calibrationist recommend assigning to one's judgment in this case? If you plan to assign credence x to the proposition that you judged on the basis of the first order evidence, then you should assign credence r to ending up with credence x in the truth and credence 1-r to ending up with credence x in the falsehood. Thus, the expected accuracy of making such a plan is:

$$(\mathrm{r})\mathbf{A}(x, \mathrm{G}) + (1 - \mathrm{r})\mathbf{A}(x, \sim\mathrm{G}). \tag{7}$$

Since we're using a strictly proper accuracy measure, this quantity is maximized when $x=r$. Thus, if you want to maximize expected accuracy, you should plan to assign credence r to the proposition that you judged on the basis of the first order evidence and credence 1-r to its negation, where r is your expected degree of reliability. This gives us a generalized version of calibrationism.

In sum, on a plausible view about what would result from planning to be steadfast, *planning* to calibrate does better than *planning* to be steadfast, even though steadfasting does better than calibrating. Planning to calibrate also does better than planning to conform to the alternative views that have been proposed in the literature.

7. Conclusion

I started this paper by pointing to some puzzling features of higher order evidence. The puzzlement boiled down to the fact that, intuitively, calibrating seems like the sort of thing we ought to be doing in cases like Aisha's. However, this judgment seems to conflict with two attractive principles about rationality.

The first principle is this: if we should think that adopting p_1 in response to evidence E will be more expectedly accurate than adopting p_2 in response to E, we can't be rationally required to adopt p_2 in response to E. This principle is at the heart of the accuracy-first program in epistemology, but it is also compelling in its own right. Why would we be required to adopt a certain credence function in response to our evidence when there's an alternative credence function that we expect to do better from the point of view of accuracy?

Here's the second principle: how confident you are now in P on the supposition that E is the same as how confident you should be in P if you go on to learn E. This principle is at the heart of Bayesian conditionalization, but it too is compelling in its own right.

The Greaves and Wallace result shows us that these two principles are intimately related. Bayesian conditionalization is the update procedure that, when certain conditions are met, maximizes expected accuracy.

Higher order evidence cases pose a prima facie problem for both of these principles. I argued, on both intuitive and formal grounds, that the update procedure that maximizes expected accuracy in cases of higher order evidence is *steadfasting*. And steadfasting is what would be warranted by straightforward conditionalization. True, we might be able to think of calibrating as a kind of conditionalizing: conditionalizing on *self-locating evidence*. But I showed that such a strategy should be rejected. Conditionalizing on this kind of evidence would require rejecting one of the assumptions ($\text{FACTIVITY}_{\text{indexed}}$) that is necessary for the magic to occur that makes conditionalizing well motivated from the point of view of accuracy.

I argued in the second part of the paper that if we think of our theorizing about rationality as *deliberating about what to believe*—that is, as making a doxastic plan—we can give an accuracy based argument for calibrating. I take this to be evidence that something like the doxastic planning framework is what underlies a deep and important notion of rationality: one that allows us to take into account, when evaluating an update procedure, our opinions concerning how successful we are likely to be at following it. It also illuminates what's right and what's wrong about the idea that the requirements of epistemic rationality can be very demanding. According to the planning framework, there is nothing inherently wrong with extremely demanding principles of rationality. So it's true that the mere fact that a principle is demanding or difficult is not a reason to rule it out. What should be ruled out in this framework, however, are principles that have the feature that planning to follow them predictably leads to worse results, from the point of view of accuracy, than planning to follow some alternative. And frequently (but not always) very demanding principles will be of this sort. Thus, the problem with steadfasting isn't simply that we can't expect to successfully conform to it. The problem with steadfasting is that there's an alternative plan (calibrating) that we should expect to lead to better results from the point of view of accuracy in cases of higher order evidence. The planning framework, which emphasizes how good it is to *make* a plan rather than how good it is to *conform* to a plan, is important not just for cognitively imperfect agents like us, but for any agent who leaves open the possibility of cognitive imperfection in the future.[27]

Appendix 1[28]

Recall that the definition of the expected accuracy of an update procedure is *the weighted average of the accuracy scores that would result from conforming to the update*

[27] For extremely helpful feedback on this work, I am grateful to audiences at Rutgers University, The University of Texas at Austin, Brandeis University's Higher Order Evidence conference, the Konstanz Reasoning Conference, Fordham University's Epistemology and Ethics Workshop, The New Insights and Directions in Religious Epistemology project at Oxford University, and the Epistemic Utility Theory project at the University of Bristol. Many thanks to Sinan Dogramaci, Branden Fitelson, Yoaav Isaacs, David Sosa, Roger White and especially Susanna Rinard for discussion and comments on earlier drafts. Some of the work for this paper was conducted at Oxford University and supported by a grant to the New Insights and Directions in Religious Epistemology project funded by the John Templeton Foundation.

[28] The arguments here are elaborated upon in greater detail in Schoenfield (forthcoming*a*).

procedure. And recall also that, for Greaves and Wallace, an agent conforms to U if she adopts $U(X_i)$ whenever *she learns* X_i. This is why we defined the expected accuracy of an update procedure as follows:

$$\mathrm{EA}^p(U) = \sum_{\mathrm{L}(X_i)\in\mathrm{L}(X)} \sum_{s\in\mathrm{L}(X_i)} p(s)^* \mathbf{A}(U(X_i), s)$$

The reason that Factivity and Partitionality are necessary for the Greaves and Wallace result is that, in their proof, Greaves and Wallace use the following formulation when calculating the expected accuracy of an update procedure:

$$\mathrm{EA}^p(U) = \sum_{X_i\in X} \sum_{s\in X_i} p(s)^* \mathbf{A}(U(X_i), s)$$

Note that the only difference between the two formulations is that, in my definition of the expected accuracy of an update procedure, we calculate the accuracy of the agent adopting $U(X_i)$ *whenever she learns* X_i, whereas in the Greaves and Wallace formulation we calculate the expected accuracy of an update procedure by imagining that the agent adopts $U(X_i)$ *whenever* X_i is true. Since Greaves and Wallace explicitly define conformity to the update procedure[29] as adopting $U(X_i)$ when X_i is *learned*, it is crucial for their result that the agent be certain that X_i is learned if and only if X_i is true. And this will be true in exactly those cases in which Factivity and Partitionality are satisfied. Here's the claim stated carefully, and the proof:

Claim:

An agent who is certain that she will learn exactly one proposition X_i from a set of propositions X will satisfy Factivity and Partitionality if and only if, for all propositions, $X_i \in X$ the agent is certain that:

$$X_i \leftrightarrow \mathrm{L}(X_i)$$

Proof:

Suppose that the agent satisfies Factivity and Partitionality.

Factivity entails certainty in the right-to-left direction of the biconditional: $\mathrm{L}(X_i) \rightarrow X_i$. For Factivity says that the agent is certain that if she learns X_i, X_i must be true.

What about the left-to-right direction? If Partitionality holds, then the agent is certain that exactly one proposition in X is true. Since the agent is certain that she will learn one proposition in X, and (due to Factivity) it will be a true proposition, she will have to learn *the* one true proposition in X. Thus, she will be certain that: $X_i \rightarrow \mathrm{L}(X_i)$.

Now for the converse: Suppose that an agent is certain that she will learn exactly one proposition from X, and that for all propostitions $X_i \in X$:

$$X_i \leftrightarrow \mathrm{L}(X_i)$$

29 They call update procedures "available acts."

Because she is certain in the right to left direction of the biconditional, the agent will satisfy FACTIVITY. But the agent must also satisfy PARTITIONALITY. For suppose that the $\mathcal{X}_i$ didn't form a partition of the agent's possibility space. Then, either the agent leaves open the possibility of more than one $\mathcal{X}_i$ being true, or she leaves open the possibility that none of the $\mathcal{X}_i$ are true. But if she left open the possibility of more than one $\mathcal{X}_i$ being true, then, by the left-to-right direction, she must leave open the possibility of learning more than one $\mathcal{X}_i$. However, we have stipulated that the agent is certain that she will learn exactly one proposition in $\mathcal{X}$. If, on the other hand, she left open the possibility of none of the $\mathcal{X}_i$ being true, then, since she is certain that exactly one proposition is going to be learned, she must leave open the possibility of learning a false proposition. This is ruled out by her certainty in the right-to-left direction.

Since Greaves and Wallace are assuming FACTIVITY and PARTITIONALITY, they can simply *define* the expected accuracy of an update procedure in response to a learning as the average accuracy scores that would result from adopting $U(\mathcal{X}_i)$ whenever $\mathcal{X}_i$ *is true*. And this, indeed, is what they do. But it's important to realize that they *wouldn't* define expected accuracy this way if they weren't assuming FACTIVITY and PARTITIONALITY. This is because, without these assumptions, the quantity:

$$\sum_{\mathcal{X}_i \in \mathcal{X}} \sum_{s \in \mathcal{X}_i} p(s)^* \mathbf{A}(U(\mathcal{X}_i)), s)$$

does not represent what they claim it's representing: the expected accuracy of the credences that would result from conforming to an update procedure.

Appendix 2

Generalized CondMax: Suppose you know that you are going to undergo a learning-experience, $\mathcal{X}$. The update procedure that maximizes expected accuracy in response to $\mathcal{X}$, relative to probability function p, is the update procedure that assigns, to each $\mathcal{X}_i$, $p(\cdot|\mathrm{L}(\mathcal{X}_i))$ where $\mathrm{L}(\mathcal{X}_i)$ is the proposition that the agent learns $\mathcal{X}_i$ upon undergoing the learning-experience.

Proof:

The purely mathematical result that can be extracted from the Greaves and Wallace paper is the following:

G&W: Take any partition of states $\mathcal{P}$: $\{\mathcal{P}_1 \ldots \mathcal{P}_n\}$ and consider the set of functions, $\mathcal{F}$, that assign members of $\mathcal{P}$ to probability functions. The member of $\mathcal{F}$, F, that maximizes this quantity:

$$\sum_{\mathcal{P}_i \in \mathcal{P}} \sum_{s \in \mathcal{P}_i} p(s)^* \mathbf{A}(F(\mathcal{P}_i), s)$$

is:

$$F(\mathcal{P}_i) = \mathrm{Cond} = p(\cdot|\mathcal{P}_i)$$

(where **A** is strictly proper).

I won't reconstruct their proof for this since it can be easily extracted from the Greaves and Wallace paper. Given this, we proceed as follows:

Recall that the expected accuracy of an update procedure, U, in response to a learning-experience $\mathcal{X}$ is defined as:

$$(\#) \qquad \sum_{L(\mathcal{X}_i)\in L(\mathcal{X})} \sum_{s\in L(\mathcal{X}_i)} p(s)^* \mathbf{A}(U(\mathcal{X}_i), s)$$

We are aiming to show is that (#) is maximized when $U(\mathcal{X}_i)$ = Cond(L($\mathcal{X}_i$)). So, suppose for reductio that this is false, that is, that there exists a function, U^*, such that:

$$\sum_{L(\mathcal{X}_i)\in L(\mathcal{X})} \sum_{s\in L(\mathcal{X}_i)} p(s)^* \mathbf{A}(U^*(\mathcal{X}_i), s) > \sum_{L(\mathcal{X}_i)\in L(\mathcal{X})} \sum_{s\in L(\mathcal{X}_i)} p(s)^* \mathbf{A}(\mathrm{Cond}(L(\mathcal{X}_i)), s)$$

Now, define μ(L($\mathcal{X}_i$)) as $U^*(\mathcal{X}_i)$.[30] It follows that:

$$\sum_{L(\mathcal{X}_i)\in L(\mathcal{X})} \sum_{s\in L(\mathcal{X}_i)} p(s)^* \mathbf{A}(\mu(L(\mathcal{X}_i)), s) > \sum_{L(\mathcal{X}_i)\in L(\mathcal{X})} \sum_{s\in L(\mathcal{X}_i)} p(s)^* \mathbf{A}(\mathrm{Cond}(L(\mathcal{X}_i)), s)$$

But this is impossible, because it follows from **G&W** that the quantity:

$$\sum_{L(\mathcal{X}_i)\in L(\mathcal{X})} \sum_{s\in L(\mathcal{X}_i)} p(s)^* \mathbf{A}(F(L(\mathcal{X}_i)), s)$$

is maximized when F = Cond(L($\mathcal{X}_i$)). So the inequality above must be false. Contradiction.

Thus, U^* does not exist: there is no update procedure that is more expectedly accurate than the update procedure that has the agent conditionalize on L($\mathcal{X}_i$) whenever she learns $\mathcal{X}_i$.

References

Arntzenius, F. (2003). "Some Problems for Conditionalization and Reflection". *Journal of Philosophy* 100(7): 356–370.

Bradley, D. (2011). "Self-Location is no Problem for Conditionalization". *Synthese* 182 (3): 393–411.

Christensen, D. (2010). "Higher Order Evidence". *Philosophy and Phenomenological Research* 81(1): 185–215.

Elga, A. (2000). "Self-Locating Belief and the Sleeping Beauty Problem". *Analysis* 60 (2): 143–147.

—— (ms). "Lucky to Be Rational".

Gibbard, A. (2008). "Rational Credence and the Value of Truth". In T. Gendler and J. Hawthorne (eds.) *Oxford Studies in Epistemology 2*. Oxford University Press.

Greaves, H. and Wallace, D. (2006). "Justifying Conditionalisation: Conditionalisation Maximizes Expected Epistemic Utility". *Mind* 115(459): 607–632.

Greco, D. (2015). "Verbal Debates in Epistemology". *American Philosophical Quarterly* 52(1): 41–55.

[30] How do we know that there is such a μ? Since there is a bijection between the $\mathcal{X}_i$ and the L($\mathcal{X}_i$) there exists an inverse of L($\mathcal{X}_i$), which we'll call "$L^-(\mathcal{X}_i)$," such that $L^-(L(\mathcal{X}_i)) = \mathcal{X}_i$. We can then let μ (L($\mathcal{X}_i$)) be U^* composed with L^-. Thus: μ(L($\mathcal{X}_i$)) = $U^*(L^-(L(\mathcal{X}_i)) = U^*(\mathcal{X}_i)$.

—— and Hedden, B. (forthcoming). “Uniqueness and Metaepistemology”. *The Journal of Philosophy*.
Halpern, J. (2005). “Sleeping Beauty Reconsidered: Conditioning and Reflection in Asynchronous Systems”. In T. Gendler and J. Hawthorne (eds.) *Oxford Studies in Epistemology 1*. Oxford University Press.
Hitchcock, C. (2004). “Beauty and Bets”. *Synthese* 139(3): 405–420.
Horowitz, S. (2013). “Immoderately Rational”. *Philosophical Studies* 167(1): 1–16.
Joyce, J. (2009). “Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief”. In F. Huber and C. Schmidt-Petri (eds.). *Degrees of Belief*. Synthese Library.
Kelly, T. (2005). “The Epistemic Significance of Disagreement”. In J. Hawthorne and T. Gendler (eds.) *Oxford Studies in Epistemology 1*. Oxford University Press.
—— (2010). “Peer Disagreement and Higher Order Evidence”. In A. Goldman and D. Whitcomb (eds.) *Social Epistemology: Essential Readings*. Oxford University Press.
Kierland, B. and Monton, B. (2005). “Minimizing Inaccuracy for Self-Locating Beliefs”. *Philosophy and Phenomenological Research* 70(2): 384–395.
Lasonen-Aarnio, M. (2014). “Higher Order Evidence and Limits of Defeat”. *Philosophy and Phenomenological Research* 88(2): 314–345.
Levinstein, B. (ms). “A Pragmatist’s Guide to Epistemic Utility”.
Meacham, C. (2008). “Sleeping Beauty and the Dynamics of De Se Beliefs”. *Philosophical Studies* 138(2): 245–269.
Moss, S. (2011). “Scoring Rules and Epistemic Compromise”. *Mind* 120(480): 1053–1069.
—— (2012). “Updating as Communication”. *Philosophy and Phenomenological Research* 85(2): 225–248.
Perry, J. (1979). “The Problem of the Essential Indexical”. *Noûs* 13(1): 3–21.
Pettigrew, R. (2016). *Accuracy and the Laws of Credence*. Oxford University Press.
Schafer, K. (2014). “Doxastic Planning and Epistemic Internalism”. *Synthese* 191(12): 2571–2591.
Schoenfield, M. (2015a). “A Dilemma for Calibrationism”. *Philosophy and Phenomenological Research* 91(2): 425–455.
—— (2015b). “Internalism without Luminosity”. *Philosophical Issues* 25(1): 252–272.
—— (forthcoming*a*). “Conditionalization Does Not, in General, Maximize Expected Accuracy”. *Mind*.
—— (forthcoming*b*). “Bridging Rationality and Accuracy”. *The Journal of Philosophy*.
Sliwa, P. and Horowitz, S. (2015). “Respecting All the Evidence”. *Philosophical Studies* 172(11): 2835–2858.
Steele, R. (forthcoming). “Anticipating Failure and Avoiding It”. *Philosophers’ Imprint*.
Titelbaum, M. (2008). “The Relevance of Self-Locating Beliefs”. *Philosophical Review* 117(4): 555–606.
—— (2015). “Rationality’s Fixed Point (Or: In Defense of Right Reason)”. In T. Gendler and J. Hawthorne (eds.) *Oxford Studies in Epistemology 5*. Oxford University Press.
Vavova, K. (forthcoming). “Irrelevant Influences”. *Philosophy and Phenomenological Research*.
Weatherson, B. (ms). “Do Judgments Screen Evidence?”.
White, R. (2009). “On Treating Oneself and Others as a Thermometer”. *Episteme* 6(3): 233–250.