Accurate Updating[1]

Ginger Schultheis
August 2022

1 Introduction

Accuracy-first epistemology aims to justify all epistemic norms by showing that they can be derived from the rational pursuit of accuracy. Take, for example, *probabilism*—the norm that credence functions should be probability functions. Accuracy-firsters say non-probabilistic credence are irrational because they're *accuracy-dominated*: For every non-probabilistic credence function, there's some probabilistic credence function that's more accurate no matter what.[2] Or take norms of *updating*, my topic in this paper. Accuracy-firsters aim to derive the rational updating rule by way of accuracy; specifically, they claim that the rational updating rule is the rule that *maximizes expected accuracy*.[3]

Externalism, put roughly, says that we do not always know what our evidence is. Though far from universally accepted, externalism is a persuasive and widely held thesis, supported by a compelling vision about the kinds of creatures we are—creatures whose information-gathering mechanisms are fallible, and whose beliefs about most subject matters are not perfectly sensitive to the facts.

Some have argued in recent years that externalists face a dilemma: Either deny that Bayesian Conditionalization is the rational update rule, thereby rejecting traditional Bayesian epistemology, or else deny that the rational update rule is the rule that maximizes expected accuracy, thereby rejecting the accuracy-first program. Call this the *Bayesian Dilemma*.

Here is roughly how the argument goes. Schoenfield (2017) has shown that following *Metaconditionalization* maximizes expected accuracy.[4] But if externalism is true, then Metaconditionalization is not Bayesian Conditionalization. Therefore, the externalist must choose between the rule that maximizes expected

[1]Thanks to David Boylan, Kevin Dorst, Matt Mandelkern, Alex Meehan, Bernhard Salow, and Pablo Zendejas Medina for helpful conversations. I am especially grateful to two people: Milo Phillips-Brown for extremely helpful feedback on earlier drafts, and to Snow Zhang for *many* conversations about this material, including conversations that led me to starting work on accuracy and updating in the first place.

[2]Joyce (2009).

[3]See Greaves & Wallace (2006) and Easwaran (2012). Not all arguments for updating norms appeal to the norm that one should maximize expected accuracy. Briggs and Pettigrew (2020) give an *accuracy-dominance* argument for Conditionalization. See also Nielsen (2021).

[4]The name of this rule is due to Das (2019).

accuracy (Metaconditionalization) and Bayesian Conditionalization.[5]

I'm not convinced by this argument. We'll see that once we make the premises fully explicit, the argument relies on assumptions that the externalist should reject. Still, I think that the Bayesian Dilemma is a genuine dilemma, and in fact I don't think that it is a special problem for externalism. I give a new argument—I call it the *continuity argument*—that does not make any assumptions that the externalist rejects. Roughly, what I show is that if you're sufficiently confident that that you would follow Metaconditionalization if you adopted Metaconditionalization, then you'll expect adopting a rule that I call *Accurate Metaconditionalization* to be more accurate than adopting Bayesian Conditionalization—and importantly, this is so *whether or not externalism is true*. The Bayesian Dilemma is a dilemma for everyone.

I'll start in §2 by introducing an accuracy-based framework for evaluating updating rules in terms of what I will call *actual inaccuracy*. In §3, I'll introduce externalism. In §4, I turn to the Bayesian Dilemma. I present an argument purporting to show that the externalist must choose between Bayesian Conditionalization and accuracy-first epistemology, and I explain why the argument does not succeed. In §5, I present the continuity argument showing that the Bayesian Dilemma is nevertheless a genuine dilemma, and that it is a dilemma for everyone. §6 concludes.

2 The Accuracy Framework: Actual Inaccuracy

Accuracy-first epistemology says that our beliefs and credal states aim at *accuracy*, or closeness to the truth; that is, our beliefs and credal states aim to avoid *inaccuracy*, or distance from the truth. We said that, according to accuracy-firsters, the rational update rule is the rule that maximizes expected accuracy. There are different ways of making that thesis precise. In this section, I argue against one common way of stating the thesis, and then offer my own. We'll start by getting the basics of the accuracy-first framework on the table.

2.1 Basics of the Accuracy Framework

For technical purposes, it is better work with measures of inaccuracy rather than measures of accuracy. An *inaccuracy measure* $\mathbf{I}$ is a function that takes a world from a set of worlds Ω, and a probability function C from a set of probability func-

[5] See Bronfman (2014), Schoenfield (2017), Das (2019), Zendejas Medina (forthcoming). Not all of these authors present their arguments as a problem for externalism. For example, Das (2019) presents the argument as a problem for accuracy-first epistemology.

tions Δ defined over $\mathcal{P}(\Omega)$, and returns a number between 0 and 1. This number represents how inaccurate C is in w. C is minimally inaccurate if it assigns 1 to all truths and 0 to all falsehoods; C is maximally inaccurate if it assigns 1 to all falsehoods and 0 to all truths.

The *expected inaccuracy* of a probability function C—relative to another probability function P—is a weighted average of C's inaccuracy in all worlds, weighted by how likely it is, according to P, that those worlds obtain. Formally:

$$\mathbb{E}_P[\mathbf{I}(C)] = \sum_{w \in \Omega} P(w) \cdot \mathbf{I}(C, w) \tag{1}$$

I will make one assumption about inaccuracy measures that will figure centrally in the arguments of §4 and §5 of this paper. Specifically:

Strict Propriety
For any two distinct probability functions P and C, $\mathbb{E}_C[\mathbf{I}(C)] < \mathbb{E}_C[\mathbf{I}(P)]$

Strict Propriety says that probabilistic credence functions expect themselves to minimize inaccuracy. Strict Propriety is often motivated by appeal to the norm of *immodesty*.[6] Though Strict Propriety is not entirely uncontroversial, it is a standard assumption in the accuracy-first literature, and I will not say anything more to justify it.

Now that we know how to measure the inaccuracy of a credence function, we turn to updating rules. I will assume that a *learning experience* can be characterized by a unique proposition—the subject's *evidence*. We define a *learning situation* as a complete specification of all learning experiences that an agent thinks she might undergo during a specific period of time—a specification of all of the propositions that the agent thinks she might learn during that time. Formally, a learning situation is an *evidence function* E that maps each world w to a proposition $\mathsf{E}(w)$, the subject's evidence in w. I will write $[\mathsf{E} = \mathsf{E}(w)]$ for the proposition that the subject's evidence is $\mathsf{E}(w)$.

$$[\mathsf{E} = \mathsf{E}(w)] = \{w' \in \Omega : \mathsf{E}(w') = \mathsf{E}(w)\} \tag{2}$$

We define an *evidential updating rule* as a function g that takes a prior probability function C, and an evidence proposition $\mathsf{E}(w)$ and returns a credence function.[7] In the next two sections of the paper, we will be talking about two updating rules. The first is Bayesian Conditionalization.

[6]See Joyce (2009), Pettigrew (2016), and Campbell-Moore & Levinstein (2021) for defenses of Strict Propriety.

[7]Not all Bayesians accept the assumption that a learning experience can be characterized by a unique proposition. Jeffrey (1965) believed that, sometimes, we undergo a learning experience,

Bayesian Conditionalization
$g_{\text{cond}}(C, \mathsf{E}(w)) = C(\cdot|\mathsf{E}(w))$

Bayesian Conditionalization says that you should respond to your evidence $\mathsf{E}(w)$ by conditioning on your evidence; for any proposition H, your new credence in H, upon receiving your new evidence, should be equal to your old credence in H conditional on your new evidence. The second rule is Metaconditionalization.

Metaconditionalization
$g_{\text{meta}}(C, \mathsf{E}(w)) = C(\cdot\ |\mathsf{E} = \mathsf{E}(w))$

Metaconditionalization says that you should respond to your evidence $\mathsf{E}(w)$ by conditioning on the proposition *that your evidence is* $\mathsf{E}(w)$.

2.2 Adopting Rules and Following Rules

I will distinguish *adopting* an updating rule from *following* an updating rule. If you *follow* a rule, then your posterior credence function is the credence function that the rule recommends. If you *adopt* an updating rule, then you intend or plan to follow the rule. Of course, in general, we can intend or plan to do things without succeeding in those doing things. Intending to follow an updating rule is no exception. We can intend or plan to follow an updating rule—in my terminology, we can *adopt* an updating rule—without following it.[8]

To see how this might happen, consider Williamson's well known case of the unmarked clock.[9] Off in the distance you catch a brief glimpse of an unmarked clock. You can tell that the hand is pointing to the upper-right quadrant of the clock, but you can't discern its exact location—your vision is good, but not perfect. What do you learn from this brief glimpse? What evidence do you gain? That—according to Williamson—depends on what the clock really reads. If the clock really reads that it is 4:05, the evidence you gain is that the time is between (say) 4:04 and 4:06. If the clock really reads 4:06, the evidence you gain is that the time is between (say) 4:05 and 4:07. Suppose that you adopt Bayesian Conditionalization as your update rule, and that the clock in fact reads 4:05. Your

but we do not learn *with certainty* that a unique proposition is true; instead, the experience tells us that a set of propositions $A_1, A_2, \ldots, A_n$ should be assigned probabilities $\alpha_1, \alpha_2, \ldots, \alpha_n$. I believe that my arguments can be recast in Jeffrey's framework, but I do not have the space to explore this question in this paper.

[8]My distinction between adopting a plan and following a plan is similar to Schoenfield (2015)'s distinction between the best plan to *follow* and the best plan to *make*. See Gallow (2021) who appeals to a related distinction between *flawless dispositions* and (potentially) *misfiring dispositions*. I discuss the relationship between Gallow's framework and my own framework in footnote 17. See also Isaacs & Russell (forthcoming).

[9]Williamson (2000).

evidence is that the time is between 4:04 and 4:06, but you mistakenly think that your evidence is that the time is between 4:05 and 4:07. As a result you misapply Bayesian Conditionalization; you condition on the wrong proposition.[10] Despite having adopted Bayesian Conditionalization as your update rule, you did not follow the rule.

The accuracy-first epistemologist says that the rational updating rule is the rule that minimizes expected inaccuracy. I said that there are different ways to make this precise. According to one common way of making it precise, the thesis is a claim about *following* updating rules (although the distinction between adopting and following is often not made explicit). At a first pass, we might understand this thesis as saying that we are rationally required to *follow* an updating rule that minimizes expected inaccuracy. But there is an immediate problem with this first-pass thesis, which others have recognized. Consider the *omniscient updating rule*, which tells you to assign credence one to all and only true propositions. The omniscient updating rule is less inaccurate than any other rule at every world, and so every probabilistic credence function expects it to uniquely minimize inaccuracy. But it would be absurd, of course, to say that we are rationally required to adopt the omniscient updating rule. To avoid this implication, theorists refine the thesis by appeal to the notion of an *available* updating rule. The refined thesis says that we're rationally required to follow an updating rule that is such that (1) following that rule is an available option and (2) following that rule minimizes expected inaccuracy among the available options.[11] Following the omniscient updating rule is not an available option and so we are not required to follow it.

To evaluate this proposal, we need to investigate the notion of availability at issue. A natural thought is that an act is available to you only if you are *able* to perform the act, and that you are able to perform an act if and only, if you tried to perform the act, you would.[12] But on this understanding, even following Bayesian Conditionalization is not always an available option. Return to the example of the unmarked clock. The clock in fact reads 4:05. Your evidence is therefore that the time is between 4:04 and 4:06. How do you update your credences? There

[10]This analysis of the case of the unmarked clock is given in Gallow (2021). Gallow uses the case as an example in which our learning dispositions can *misfire*. I discuss the relationship between my framework and Gallow's framework in footnote 18.

[11]This is roughly how Greaves & Wallace (2006), Schoenfield (2017), and Das (2019) understand it.

[12]For defenses of the view that the scope of our options is limited to the scope of our abilities, see Richard Jeffrey (1965), Jeffrey (1992), Lewis (1981), Hedden (2012), and Koon (2020). For example, Jeffrey (1965) regards options as propositions and writes, 'An act is then a proposition which is within the agent's power to make true if he pleases.'

are two cases. In the first case, you correctly identify your evidence, and as a result, you condition on your evidence. In this case, it is true if you tried to follow Bayesian Conditionalization, you would. In the second case, you mistakenly take your evidence to be that the time is between 4:05 and 4:07, and as a result, you condition on the wrong proposition. In this case, it is *not* true that if you tried to follow Bayesian Conditionalization, then you would, and so it is not true that you are able to follow Bayesian Conditionalization. Of course, one might object to this account of ability. Rather than wade any further into this debate, I will simply observe that *however* we define availability, if we state the accuracy-first thesis in terms of following, we'll be taking for granted that if you adopt an updating rule, you will follow it; we'll be ignoring possibilities in which you do not succeed in following your updating rule because you mistake your evidence. But cases like the unmarked clock suggest that cases like this are commonplace; we should take them into account. In light of this, I suggest that we understand that the accuracy-first thesis as a thesis about which updating rule we are rationally required to *adopt*. To that end, we need to say how to evaluate the inaccuracy of adopting an updating rule.

2.3 Actual Inaccuracy

I propose to measure the inaccuracy of adopting an updating rule in terms of what I will call *actual inaccuracy*.[13] Roughly, the actual inaccuracy of adopting an updating rule g in a world w is the inaccuracy, in w, of the credence function you would have if you adopted g in w. To give a more precise definition, I need to introduce *credal selection functions*.

A credal selection function is a function f that takes an evidential updating rule g and a world w, and returns a credence function $f(g, w)$, the credence function that the subject would have if she were to adopt the rule g in world w.[14] Of course any number of factors might play a role in determining what credence function a given subject would have if she were to adopt a certain updating rule. To keep things manageable, I am going to make some assumptions about how we are disposed to change our credal states, assumptions that simplify the presentation of the main arguments of this paper, but are inessential for the proofs

[13]This term comes from Andrew Bacon's notion of *actual value*. See Bacon (2022).

[14]Credal selection functions can be defined in terms of *Stalnakerian* selection functions. A Stalnakerian selection function h—used in Stalnaker's (1968) semantics for conditionals—is a function that takes a proposition A and a world w and returns another world $h(A, w)$—intuitively, the world that would have obtained if A had been true in w. Then where *Adopt-g* is the proposition that the subject adopts updating rule g, we can define $f(g, w)$ as the credence function you have in $h(\textit{Adopt-}g, w)$.

that I give later in this paper.

Return to the example of the unmarked clock. Suppose you adopt Bayesian Conditionalization. In fact, the clock reads 4:05 and so your evidence is that the time is between 4:04 and 4:06. How do you update your credences? There are, as before, two cases. In one case, you correctly identify your evidence: to use the terminology that I will from now on adopt, you *guess* correctly that your evidence is that the time is between 4:04 and 4:06. In this case, the conditional *if you adopted Bayesian Conditionalization, then you would follow Bayesian Conditionalization* is true of you. In the second case, you guess incorrectly that your evidence is that the time is between 4:05 and 4:07. In this case, the conditional *if you adopted Bayesian Conditionalization, then you would follow Bayesian Conditionalization* is false. Instead, the following conditional is true: *if you adopted Bayesian Conditionalization, then the credence function you would have is the credence function that results from applying Bayesian Conditionalization to the proposition that the time is between 4:05 and 4:07.* I will assume that these are the only two cases. Either you guess correctly and condition on the right proposition; or else you guess wrong and condition on the wrong proposition.

To make this more precise, fix a set of worlds Ω and an evidence function E defined on Ω. We will let G^{E} be a *guess function* defined on Ω. This is a function that takes each world w to a proposition $\mathsf{G}^{\mathsf{E}}(w)$: the subject's guess about what her evidence is in w.[15] Then I will characterize the subject's credal selection function as follows. Where f_{C,G^E} is the credal selection for a subject whose credence function is C and whose guess function is G^{E}, then for any evidential updating rule g:[16]

$$f_{C,\mathsf{G}^{\mathsf{E}}}(g, w) = g(C, \mathsf{G}^{\mathsf{E}}(w)) \tag{3}$$

For example, consider Bayesian Conditionalization. (3) gives us:

$$f_{C,\mathsf{G}^{\mathsf{E}}}(g_{\text{cond}}, w) = g_{\text{cond}}(C, \mathsf{G}^{\mathsf{E}}(w)) = C(\cdot|\mathsf{G}^{\mathsf{E}}(w)) \tag{4}$$

The credence function you would have if you adopted Bayesian Conditionalization is the result of conditioning your prior on your guess about what your evidence is. Likewise, consider Metaconditionalization. (3) says that the credence

[15] Isaacs & Russell (forthcoming) also use the term 'guess function'. Note, however, that they use the term differently from how I am using it here. In particular, their guess functions are used to model guesses about which *world* you are in. (In their framework, worlds are *coarse*—they settle some questions, but not all.) There are many interesting connections between my framework and the framework used in Isaacs & Russell, but I do not have the space to address them here. See Schultheis (ms).

[16] Here I assume that $\mathsf{G}^{\mathsf{E}}(w) = \mathsf{E}(w')$ for some $w' \in \Omega$.

function you would have if you adopted Metaconditionalization is the result of conditioning your prior on the proposition that your evidence is $\mathsf{G}^{\mathsf{E}}(w)$, your guess about what your evidence is in w. Formally:

$$f_{C,\mathsf{G}^{\mathsf{E}}}(g_{\text{meta}}, w) = g_{\text{meta}}(C, \mathsf{G}^{\mathsf{E}}(w)) = C(\cdot|\mathsf{E} = \mathsf{G}^{\mathsf{E}}(w)) \tag{5}$$

We'll now use (3) to give a more precise definition of the the actual inaccuracy of adopting an evidential updating rule. Let E be any evidence function, G^{E} any guess function, and C any prior. For any evidential updating rule g, and any world w, we define $V_{C,\mathsf{G}^{\mathsf{E}}}(g, w)$: the actual inaccuracy of adopting g in w for a subject with prior C and guess function G^{E}.

Actual Inaccuracy
$V_{C,\mathsf{G}^{\mathsf{E}}}(g, w) = \mathbf{I}(f_{C,\mathsf{G}^{\mathsf{E}}}(g, w), w) = \mathbf{I}(g(C, \mathsf{G}^{\mathsf{E}}(w)), w)$

Here's what this says. Consider a subject in learning situation E whose guess function is G^{E}. The actual inaccuracy of adopting an evidential updating rule g, in a world w, for this subject, is the inaccuracy, in w, of the credence function the subject would have, given her guess function, if she adopted g in w.[17]

The *expected actual inaccuracy* of adopting an evidential updating rule g, relative to C, is a weighted average of the actual inaccuracy of adopting g in all possible states of the world, weighted by how likely it is, according to C, that those states obtain. Formally:

$$\sum_{w \in \Omega} C(w) \cdot V_{C,\mathsf{G}^{\mathsf{E}}}(g, w) = \sum_{w \in \Omega} C(w) \cdot \mathbf{I}(f_{C,\mathsf{G}^{\mathsf{E}}}(g, w), w) \tag{6}$$

Return to the accuracy-first thesis that the rational updating rule is the rule that does best in terms of accuracy. I have argued that this claim is best understood as a claim about which updating rule we should *adopt*. We can now make this claim more precise using the notion of actual inaccuracy. I propose to formulate the accuracy-first thesis, which I call *Accuracy-First Updating*, as follows.[18]

[17]Note that the actual inaccuracy of adopting g in w is not always the inaccuracy of your credence function in w. Suppose you do not adopt g in w. Then the actual inaccuracy of adopting g in w is the inaccuracy, in w, of the credence function you *would* have if you *had* adopted g in w.

[18]It is worth taking a moment to compare my framework with the framework in Gallow (2021). Gallow is also concerned to account for subjects who are not sure that they will respond correctly to their evidence. There are nevertheless important differences between our frameworks and between the questions that we are interested in. Gallow is primarily interested in the question: *Which (potentially misfiring) learning disposition is best?* A learning disposition in Gallow's framework is not the same thing as a credal selection function in my framework. Instead, a learning disposition is closer to a function $f(g, \cdot)$, where f is a credal selection function and g is an updating rule. This is a function that takes a world w, and returns the credence function you

Accuracy-First Updating
You are rationally required to adopt an evidential updating rule that minimizes expected actual inaccuracy.

Let's turn now to epistemic externalism.

3 Externalism

To characterize externalism, we need to first characterize internalism. Internalism says, roughly, that for certain special propositions, when those propositions are true, we have a special kind of *access* to their truth. Let's say that you have access to a proposition if and only, whenever it is true, your evidence entails that it is true. Then internalism says that, for certain special propositions, whenever those propositions are true, your evidence entails that they are true. There are different brands of internalism, depending on what kinds of propositions are taken to be special. According to some, the special propositions are propositions about our own minds, such as the proposition that I am in pain. These internalists say that, whenever I am in pain, my evidence entails that I am in pain—I can always tell that I am in pain by carefully attending to this evidence, my own experiences. In this paper, we will be mainly interested in one form of internalism—*evidence internalism*. On this view, propositions *about what our evidence is* are special propositions in the sense that whenever they're true, our evidence entails that they are true.

Evidence Internalism
If your evidence is the proposition $\mathsf{E}(w)$, then your evidence entails *that* your evidence is $\mathsf{E}(w)$.

Let *evidence externalism* be the denial of evidence internalism. More precisely:

Evidence Externalism
Sometimes, your evidence is some proposition $\mathsf{E}(w)$, but your evidence does not entail that your evidence is $\mathsf{E}(w)$.

Why accept evidence externalism? One standard argument appeals to our fallibility. The externalist says that all of our information-gathering mechanisms

will have in w, given that you have adopted the updating rule g. Translated into my framework, Gallow's question becomes: *Which $f(g, \cdot)$ is best to have?* That is not my question. My question is: *Holding fixed your credal selection function, which updating rule is it best for you to adopt?* Put differently: *Holding fixed the conditional facts—the facts about which credence function you would have if you adopted this rule or that rule—which rule is it best for you to adopt?*

are fallible. Now, it is no surprise that our mechanisms specialized for detecting the state of our external environment—such as whether it is raining, or whether there is a computer on my desk—can lead us astray. What is controversial about externalism is its insistence that what is true of these propositions about my external environment is true of nearly all propositions, including the proposition that I am in pain or that I feel cold. The externalist says that, sometimes, I am feeling cold, but my mechanisms specialized for detecting feelings of coldness misfire, telling me that I am not feeling cold.

The externalist asks us to consider a case in which your information-gathering mechanisms have misfired. As a matter of fact, I'm feeling cold, but my mechanisms specialized for detecting feelings of coldness misfire, telling me that I'm not feeling cold. Since it is false that I'm not feeling cold, it is not part of my *evidence* that I'm not feeling cold. But I have no reason to believe that anything is amiss—it is not part of my evidence *that* it is not part of my evidence that I'm not feeling cold. Evidence externalism holds.[19]

4 The Bayesian Dilemma and the Externalist Reply

In the introduction I said that some have argued that externalists face a dilemma, the *Bayesian Dilemma*: Either deny that we are rationally required to adopt Bayesian Conditionalization as our update rule or else deny that the rational update rule is the rule that maximizes expected accuracy, thereby rejecting the accuracy-first program. In this section, I present a core piece of that argument, Schoenfield's result that you can expect following Metaconditionalization to be more accurate than following any other updating rule. But as we'll see, this result cannot do the work that others have thought it can. It doesn't follow from Schoenfield's result that you expect *adopting* Metaconditionalization to be more accurate than adopting Bayesian Conditionalization, and I have argued that that it is adopting, not following, that the accuracy-first updating thesis should concern.

Let's begin by stating Schoenfield's result.

Theorem 1
Let E be any learning situation and let C be any probability function such that $C(\mathsf{E} = \mathsf{E}(w)) > 0$ for all $w \in \Omega$. Let g be any evidential updating rule that disagrees with Metaconditionalization for C at some world $w \in \Omega$ in E. Then:

[19]Versions of this argument can be found in McDowell (1982, 2011), Williamson (2000), Weatherson (2011), Salow (2019).

$$\sum_{w\in\Omega} C(w)\cdot \mathbf{I}(g_{\text{meta}}(C,\mathsf{E}(w)) < \sum_{w\in\Omega} C(w)\cdot \mathbf{I}(g(C,\mathsf{E}(w))$$

Here is what Theorem 1 says. Consider any evidential updating rule g that disagrees with Metaconditionalization in learning situation E. Consider any subject who leaves open worlds where they disagree. (We assume $C(\mathsf{E} = \mathsf{E}(w)) > 0$ for all $w \in \Omega$ to guarantee that our subject does leave open some such worlds.) Then, Theorem 1 says, the subject will expect the recommendation of Metaconditionalization to be strictly less inaccurate than the recommendation of g in that learning situation.

But, as Schoenfield and others observe, if evidence externalism is true, Metaconditionalization is not Bayesian Conditionalization. Remember, Conditionalization says that you should respond to your evidence $\mathsf{E}(w)$ by conditioning on $\mathsf{E}(w)$. Metaconditionalization says that you should respond to $\mathsf{E}(w)$ by conditioning on the proposition that your evidence is $\mathsf{E}(w)$, the proposition $[\mathsf{E} = \mathsf{E}(w)]$. If evidence externalism is true, then $\mathsf{E}(w)$ is not the same proposition as $[\mathsf{E} = \mathsf{E}(w)]$. In particular, $\mathsf{E}(w)$ may not entail the proposition $[\mathsf{E} = \mathsf{E}(w)]$, and when this happens, Metaconditionalization and Bayesian Conditionalization may disagree. Thus, if evidence externalism is true, then Theorem 1 entails that the subject will expect the recommendation of Metaconditionalization to be less inaccurate than the recommendation of Bayesian Conditionalization. In other words, if evidence externalism is true, then the subject will expect following Metaconditionalization to be less inaccurate than following Bayesian Conditionalization.

But it doesn't follow from Theorem 1 that she expects *adopting*—intending or planning to follow—Metaconditionalization to be less inaccurate than adopting Bayesian Conditionalization. That would follow from Theorem 1 only if the subject were *sure* that she would follow Metaconditionalization if she adopted Metaconditionalization.

To see this, let G^{E} be the subject's guess function. Let *Guess Right* be the proposition that the subject's guess about her evidence in learning situation E is right. Formally:

$$\textit{Guess Right} = \{w \in \Omega : \mathsf{G}^{\mathsf{E}}(w) = \mathsf{E}(w)\} \tag{7}$$

Let *Guess Wrong* be the proposition that the subject's guess about her evidence in E is not right. Formally:

$$\textit{Guess Wrong} = \{w \in \Omega : \mathsf{G}^{\mathsf{E}}(w) \neq \mathsf{E}(w)\} \tag{8}$$

If $C(Guess\ Right) = 1$, then I will say that the subject is *infallible*; otherwise, I will say that the subject is *fallible*.

Suppose our subject is infallible. She is sure that she will correctly identify her evidence. Then, for any learning situation E, and any evidential updating rule g:

$$\text{If } C(w) > 0, \text{ then } f_{C,\mathsf{G}^{\mathsf{E}}}(g, w) = g(C, \mathsf{G}^{\mathsf{E}}(w)) = g(C, \mathsf{E}(w)) \tag{9}$$

The subject is sure that she would follow whatever evidential updating rule that she adopted. In that case, Theorem 1 entails:

$$\sum_{w \in \Omega} C(w) \cdot \mathbf{I}(f_{C,\mathsf{G}^{\mathsf{E}}}(g_{\text{meta}}, w), w) < \sum_{w \in \Omega} C(w) \cdot \mathbf{I}(f_{C,\mathsf{G}^{\mathsf{E}}}(g, w), w) \tag{10}$$

(10) says that the subject expects adopting Metaconditionalization to be strictly less inaccurate than adopting any other evidential updating rule. If externalism is true, (10) entails that the subject expects adopting Metaconditionalization to be strictly less inaccurate than adopting Bayesian Conditionalization.

But we can't simply *assume* that the subject is infallible—that she is sure that she will correctly identify her evidence. To assume that the subject is infallible, it seems to me, comes close to begging the question against evidence externalism. Remember, the externalist says that all of our information-gathering mechanisms are fallible. Sometimes, she says, I feel cold, but my mechanisms specialized for detecting whether or not I feel cold misfire; as a result, I falsely believe that I do not feel cold. As we have seen, this fallibility also applies to propositions *about our evidence*. Return to the case of the unmarked clock. The evidence you gain is that the time is between 4:04 and 4:06. But you mistakenly think that your evidence is some other proposition—that the time is between 4:05 and 4:07. If you're an externalist, you know this about yourself; you know that your information-gathering mechanisms are fallible, and that, as a result, you will sometimes mistake your evidence.

Let me summarize. If evidence externalism is true, then Theorem 1 tells us that you will expect following Metaconditionalization to be less inaccurate than following any other evidential updating rule. It doesn't follow, however, that you expect *adopting* Metaconditionalization to be less inaccurate than adopting any other rule. In particular, it doesn't follow that you expect adopting Metaconditionalization to be less inaccurate than adopting Bayesian Conditionalization. That would follow only if we knew that you were infallible, but we cannot, on pain of begging the question against the externalist, simply assume that you are infallible. So we have not shown that the externalist must choose between

Accuracy-First Updating and Bayesian Conditionalization. We have not shown that the Bayesian Dilemma is a genuine dilemma.[20]

5 The Bayesian Dilemma Reconsidered

In this section, I show that we can establish the Bayesian Dilemma without the assumption of infallibility. I give a new argument—I call it the *continuity argument*—showing that if you are sufficiently confident, but not necessarily certain, that you will correctly identify your evidence, then adopting a rule that I'll call *Accurate Metaconditionalization* will have less expected inaccuracy than adopting Bayesian Conditionalization. I'll begin by saying what Accurate Metaconditionalization is, and then I'll present the continuity argument. This result shows that the Bayesian Dilemma is a dilemma for everyone. I'll then consider one way in which the continuity argument might be thought to pose an additional, distinctive threat to externalism, but I'll argue, the *core* commitments of externalism are not under threat.

5.1 The Continuity Argument

Metaconditionalization said that you should respond to your evidence $\mathsf{E}(w)$ by conditioning on the proposition that your evidence is $\mathsf{E}(w)$. Accurate Metaconditionalization says that you should respond to your evidence $\mathsf{E}(w)$ by conditioning on the proposition that your evidence is $\mathsf{E}(w)$ *and* that you have guessed right.

[20] Here I state the Bayesian Dilemma in terms of *adopting* an updating rule because I prefer to state the accuracy-first thesis as a thesis about rule adoption, not a thesis about rule following. As I mentioned in §2, many theorists (implicitly) take the accuracy-first thesis to be a thesis about *following*. For these theorists, the Bayesian Dilemma is a choice between (a) the claim that we're required to *follow* Bayesian Conditionalization and (b) the claim that we're required to follow a rule that minimizes expected inaccuracy. The argument for this version of the Bayesian Dilemma runs as follows. Following Metaconditionalization is an available option, and following Metaconditionalization minimizes expected inaccuracy among the available options. Therefore, if accuracy-first epistemology is true, we're required to follow Metaconditionalization. But if externalism is true, Metaconditionalization is not Bayesian Conditionalization. So the externalist must choose between accuracy-first epistemology and Bayesian Conditionalization. I don't think the externalist should be persuaded by this version of the argument, either. In particular, they should deny that following Metaconditionalization is (always) an available option. Earlier I said that a standard constraint on option availability is that an act is available only if you are *able* to perform the act. But the externalist simply must deny that we are always able to follow Metaconditionalization. Remember, the externalist says that my information-gathering mechanisms are fallible, and that as a result, my beliefs about what evidence I have are not perfectly sensitive to the facts about what evidence I have. Importantly, for the externalist, there's nothing I can do to change this—no amount of careful attention to my evidence will insure me against error. I cannot help but form false about what my evidence is in at least some learning situations; that is, I am not able to follow Metaconditionalization in all learning situations.

More precisely:

Accurate Metaconditionalization
Where C is any prior such that $C(\mathsf{E} = \mathsf{E}(w)|\textit{Guess Right}) > 0$ for all $w \in \Omega$:
$g_{\text{acc-meta}}(C, \mathsf{E}(w)) = C(\cdot|\textit{Guess Right} \wedge \mathsf{E} = \mathsf{E}(w))$

Remember, we are assuming that the credence function you would have if you adopted an evidential updating rule g in w is the result of applying g to your guess in w, $\mathsf{G}^{\mathsf{E}}(w)$. Formally:

$$f_{C,\mathsf{G}^{\mathsf{E}}}(g, w) = g(C, \mathsf{G}^{\mathsf{E}}(w)) \tag{11}$$

Suppose that $C(\mathsf{E} = \mathsf{E}(w)|\textit{Guess Right}) > 0$ for all $w \in \Omega$. Then (11) implies:

$$f_{C,\mathsf{G}^{\mathsf{E}}}(g_{\text{acc-meta}}, w) = g_{\text{acc-meta}}(C, \mathsf{G}^{\mathsf{E}}(w)) = C(\cdot|\textit{Guess Right} \wedge \mathsf{E} = \mathsf{G}^{\mathsf{E}}(w)) \tag{12}$$

(12) says that the credence function you would have if you adopted Accurate Metaconditionalization is the result of conditioning your prior on the proposition that your evidence is $\mathsf{G}^{\mathsf{E}}(w)$, your guess about what your evidence is in w, *and that you have guessed right.*

I am going to show that, for a wide class of fallible subjects, if the subject is sufficiently confident in *Guess Right*, and she leaves open worlds where Bayesian Conditionalization and Accurate Metaconditionalization disagree, then she will expect Accurate Metaconditionalization to be strictly less inaccurate than Bayesian Conditionalization. Let me say roughly how the continuity argument is going to go. I will begin by showing that we can state the expected actual inaccuracy of adopting an updating rule as a function of your credence x in the proposition *Guess Right*. In particular, we can state the expected actual inaccuracy of adopting Accurate Metaconditionalization as a function of x, and we can state the expected actual inaccuracy of adopting Bayesian Conditionalization as a function of x. Importantly, both functions are continuous functions of x. We know that when $x = 1$ and the subject is infallible, adopting Bayesian Conditionalization has greater expected actual inaccuracy than adopting Accurate Metaconditionalization. Since both functions are continuous, it follows there is some $\gamma > 0$ such that if $x > 1 - \gamma$, then adopting Bayesian Conditionalization has greater expected actual inaccuracy than adopting Accurate Metaconditionalization.

Let's now turn to the details. To begin, I am going to introduce and define a new class of functions. I will call them *probability extension functions*. We can think of a probability extension function as a specification of the conditional credences of some hypothetical subject, conditional on each member of the partition

$\{Guess\ Right, Guess\ Wrong\}$ that the subject leaves open. We then feed the probability extension function a possible credence x in *Guess Right* (a real number between 0 and 1) and the function returns a (complete) probability function—the probability function determined by the conditional credence specifications, together with x.

To make this more precise, fix a set of worlds Ω. Let E be any evidence function, and let G^{E} be any guess function. Remember that Δ is the set of probability functions over $\mathcal{P}(\Omega)$. We define Δ_{Right} as follows.

$$\Delta_{\text{Right}} = \{P_R : P_R \in \Delta \text{ and } P_R(Guess\ Right) = 1\} \tag{13}$$

And we define Δ_{Wrong} in a similar way.

$$\Delta_{\text{Wrong}} = \{P_W : P_W \in \Delta \text{ and } P_W(Guess\ Wrong) = 1\} \tag{14}$$

For each pair $\langle P_R, P_W \rangle$ consisting of a $P_R \in \Delta_{\text{Right}}$ and a $P_W \in \Delta_{\text{Wrong}}$, we define a probability extension function $\lambda_{\langle P_R, P_W \rangle}$ as a function that takes a real number x between 0 and 1 and returns a probability function $\lambda_{\langle P_R, P_W \rangle}(x)$ over $\mathcal{P}(\Omega)$ defined as follows.

$$\lambda_{\langle P_R, P_W \rangle}(x)(\cdot) = P_R(\cdot)x + P_W(\cdot)(1 - x) \tag{15}$$

Note that for any probability function C over $\mathcal{P}(\Omega)$, there is some probability extension function $\lambda_{\langle P_R, P_W \rangle}$ such that:[21]

$$C(\cdot) = \lambda_{\langle P_R, P_W \rangle}(C(Guess\ Right)) \tag{16}$$

In this way, probability extension functions allow us to specify a subject's credence in any proposition as a function of her credence in *Guess Right*.

We can also use probability extension functions to specify the expected actual inaccuracy of adopting an updating rule, for some subject, as a function of her credence in *Guess Right*. To see this, fix a learning situation E, a guess function G^{E}, a prior probability function C, and an evidential updating rule g. Each probability extension function $\lambda_{\langle P_R, P_W \rangle}$ determines a function that takes a credence x in *Guess Right* and returns the expectation, relative to $\lambda_{\langle P_R, P_W \rangle}(x)$, of the actual inaccuracy of adopting rule g *with respect to prior* C, in learning situation E, given

[21] If $C(Guess\ Right) > 0$ and $C(Guess\ Wrong) > 0$, then let $P_R(\cdot) = C(\cdot|Guess\ Right)$ and let $P_W(\cdot) = C(\cdot|Guess\ Wrong)$. If $C(Guess\ Wrong) = 1$, then let P_R be any probability function in Δ_{right}, and let $P_W(\cdot) = C(\cdot)$. If $C(Guess\ Right) = 1$, let P_W be any probability function in Δ_{wrong} and let $P_R(\cdot) = C(\cdot)$.

guess function G^{E}. For example, consider:

$$\mathbb{E}_{\lambda_{\langle P_R, P_W \rangle}(x)}[\mathbf{I}(f_{C,\mathsf{G}^{\mathsf{E}}}(g_{\text{meta}}))] = \sum_{w \in \Omega} \lambda_{\langle P_R, P_W \rangle}(x)(w) \cdot \mathbf{I}[g_{\text{meta}}(C, \mathsf{G}^{\mathsf{E}}(w)), w] \tag{17}$$

This is a function that takes a possible credence x in *Guess Right* and returns the expectation, relative to $\lambda_{\langle P_R, P_W \rangle}(x)$, of adopting Metaconditionalization with respect to prior C in learning situation E, given guess function G^{E}. Similarly, consider:

$$\mathbb{E}_{\lambda_{\langle P_R, P_W \rangle}(x)}[\mathbf{I}(f_{C,\mathsf{G}^{\mathsf{E}}}(g_{\text{cond}}))] = \sum_{w \in \Omega} \lambda_{\langle P_R, P_W \rangle}(x)(w) \cdot \mathbf{I}[g_{\text{cond}}(C, \mathsf{G}^{\mathsf{E}}(w)), w] \tag{18}$$

This is a function that takes a possible credence x in *Guess Right* and returns the expectation, relative to $\lambda_{\langle P_R, P_W \rangle}(x)$, of the actual inaccuracy of adopting Bayesian Conditionalization with respect to prior C in learning situation E, given guess function G^{E}.

We are now ready to state and prove our main theorem.

Theorem 2
Let E be any evidence function, G^{E} any guess function, and P any prior such that $P(\mathsf{E} = \mathsf{E}(w) | \textit{Guess Right}) > 0$ for all $w \in \Omega$. Let g be any updating rule that disagrees with Accurate Metaconditionalization for P at some world in learning situation E. Then there is a $\gamma > 0$ such that, if $P(\textit{Guess Right}) > 1 - \gamma$, then:

$$\sum_{w \in \Omega} P(w) \cdot \mathbf{I}[f_{P,\mathsf{G}^{\mathsf{E}}}(g_{\text{acc-meta}}, w), w] < \sum_{w \in \Omega} P(w) \cdot \mathbf{I}[f_{P,\mathsf{G}^{\mathsf{E}}}(g, w), w]$$

Consider any evidential updating rule g that disagrees with Accurate Metaconditionalization at some world in learning situation E. Consider any subject such that, for any world $w \in \Omega$, the subject leaves open $[\mathsf{E} = \mathsf{E}(w)]$, conditional on *Guess Right*. Then, Theorem 2 says, if the subject is sufficiently confident that *Guess Right* is true, then she will expect adopting Accurate Metaconditionalization to be strictly less inaccurate than adopting g.

The proof of Theorem 1 relies on two lemmas. First we have Lemma 1.

Lemma 1
Let E be any evidence function, G^{E} any guess function, and C any infallible prior such that $C(\mathsf{E} = \mathsf{E}(w)) > 0$ for all $w \in \Omega$. Let P be any (possibly fallible) prior, and let g be any updating rule such that: $g(P, \mathsf{E}(w)) \neq$

$g_{\text{meta}}(C, \mathsf{E}(w))$ for some $w \in \Omega$. Then:

$$\sum_{w \in \Omega} C(w) \cdot \mathbf{I}[f_{C,\mathsf{G}^{\mathsf{E}}}(g_{\text{meta}}, w), w] < \sum_{w \in \Omega} C(w) \cdot \mathbf{I}[f_{P,\mathsf{G}^{\mathsf{E}}}(g, w), w]$$

Here's roughly what this says. Suppose you are infallible; that is, you are sure that *Guess Right* is true. Consider any (possibly fallible) prior P and any evidential updating rule g such that, for some worlds, following rule g *with respect to* P disagrees with following Metacondionalization *with respect to your infallible prior*. Then will expect adopting Metaconditionalization *with respect to your infallible prior* to have strictly lower inaccuracy than following g with respect to the possibly fallible prior P. I leave the proof of Lemma 1 to an appendix.

The second lemma is Lemma 2.

Lemma 2
Let E be any learning situation, G^{E} any guess function, g any evidential updating rule, $\lambda_{\langle P_R, P_W \rangle}$ any probability extension function, and C any probability function. Then:

$$\sum_{w \in \Omega} \lambda_{\langle P_R, P_W \rangle}(x)(w) \cdot \mathbf{I}[f_{C,\mathsf{G}^{\mathsf{E}}}(g, w), w]$$

is a continuous function of x.

The proof of Lemma 2 is straightforward. Observe that:

$$\sum_{w \in \Omega} \lambda_{\langle P_R, P_W \rangle}(x)(w) \cdot \mathbf{I}[f_{C,\mathsf{G}^{\mathsf{E}}}(g, w), w] \tag{19}$$

is a sum of terms of the form:

$$\lambda_{\langle P_R, P_W \rangle}(x)(w) \cdot \mathbf{I}(f_{C,\mathsf{G}^{\mathsf{E}}}(g, w), w) \tag{20}$$

where $\mathbf{I}(f_{C,\mathsf{G}^{\mathsf{E}}}(g, w), w)$ is a constant and $\lambda_{\langle P_R, P_W \rangle}(x)(w)$ is a continuous function of x.[22] Thus, (19) is a linear combination of continuous functions of x, and so is itself a continuous function of x.[23]

We will now use Lemma 1 and Lemma 2 to complete the proof of Theorem 2. Let E be any evidence function, and let G^{E} be any guess function. Let g be

[22] $\lambda_{\langle P_R, P_W \rangle}(x)(w) = P_R(w)x + P_W(w)(1 - x)$ is a polynomial and so is continuous everywhere.

[23] Note that the fact (19) is a continuous function of x does not depend on the assumption that the inaccuracy measure $\mathbf{I}$ is itself continuous. Like the proofs of Greaves & Wallace (2006), Schoenfield (2017), and Das (2019), my argument makes only one assumption about $\mathbf{I}$—namely, that $\mathbf{I}$ is strictly proper. Strict Propriety is defined in §2.

any evidential updating rule such that $g(P, \mathsf{E}(w)) \neq g_{\text{acc-meta}}(P, \mathsf{E}(w))$ for some $w \in \Omega$. Consider any fallible prior P such that $P(\textit{Guess Right}) > 0$, and $P(\mathsf{E} = \mathsf{E}(w) | \textit{Guess Right}) > 0$ for all $w \in \Omega$. Let $C(\cdot) = P(\cdot | \textit{Guess Right})$.

We know that $g(P, \mathsf{E}(w)) \neq g_{\text{acc-meta}}(P, \mathsf{E}(w))$ for some $w \in \Omega$. We also know that $g_{\text{acc-meta}}(P, \mathsf{E}(w)) = g_{\text{meta}}(C, \mathsf{E}(w))$. It follows that:

$$g(P, \mathsf{E}(w)) \neq g_{\text{meta}}(C, \mathsf{E}(w)) \text{ for some } w \in \Omega \tag{21}$$

By Lemma 1, (21) entails:

$$\sum_{w \in \Omega} C(w) \cdot \mathbf{I}[f_{C,\mathsf{G}^{\mathsf{E}}}(g_{\text{meta}}, w), w] < \sum_{w \in \Omega} C(w) \cdot \mathbf{I}[f_{P,\mathsf{G}^{\mathsf{E}}}(g, w), w] \tag{22}$$

We know there is some probability extension function $\lambda_{\langle P_R, P_W \rangle}$ such that $P(\cdot) = \lambda_{\langle P_R, P_W \rangle}(P(\textit{Guess Right}))(\cdot)$. It is also easy to show:[24]

$$\lambda_{\langle P_R, P_W \rangle}(1)(\cdot) = P(\cdot | \textit{Guess Right}) = C(\cdot) \tag{28}$$

(22) and (28) together entail:

$$\sum_{w \in \Omega} \lambda_{\langle P_R, P_W \rangle}(1)(w) \cdot \mathbf{I}[f_{C,\mathsf{G}^{\mathsf{E}}}(g_{\text{meta}}, w), w] < \sum_{w \in \Omega} \lambda_{\langle P_R, P_W \rangle}(1)(w) \cdot \mathbf{I}[f_{P,\mathsf{G}^{\mathsf{E}}}(g, w), w] \tag{29}$$

Lemma (2) says that

$$\sum_{w \in \Omega} \lambda_{\langle P_R, P_W \rangle}(x)(w) \cdot \mathbf{I}[f_{C,\mathsf{G}^{\mathsf{E}}}(g_{\text{meta}}, w), w] \tag{30}$$

and

$$\sum_{w \in \Omega} \lambda_{\langle P_R, P_W \rangle}(x)(w) \cdot \mathbf{I}[f_{P,\mathsf{G}^{\mathsf{E}}}(g, w), w] \tag{31}$$

are both continuous functions of x. This, together with (29), implies:

[24] *Proof.*

$$\begin{aligned}
P(H | \textit{Guess Right}) &= \frac{P(H \wedge \textit{Guess Right})}{P(\textit{Guess Right})} && (23)\\
&= \frac{P_R(H)x}{P_R(\textit{Guess Right})x} && (24)\\
&= \frac{P_R(H)}{P_R(\textit{Guess Right})} && (25)\\
&= P_R(H) && (26)\\
&= \lambda_{\langle P_R, P_W \rangle}(1)(H) && (27)
\end{aligned}$$

$$\text{There's a } \delta > 0 \text{ such that, if } x > 1 - \delta \text{, then} \tag{32}$$
$$\sum_{w \in \Omega} \lambda_{\langle P_R, P_W \rangle}(x)(w) \cdot \mathbf{I}[f_{C,\mathsf{G}^\mathsf{E}}(g_{\text{meta}}, w), w] < \sum_{w \in \Omega} \lambda_{\langle P_R, P_W \rangle}(x)(w) \cdot \mathbf{I}[f_{P,\mathsf{G}^\mathsf{E}}(g, w), w]$$

Since $P(\cdot) = \lambda_{\langle P_R, P_W \rangle}(P(\textit{Guess Right}))$, (32) entails:

$$\text{There's a } \delta > 0 \text{ such that, if } P(\textit{Guess Right}) > 1 - \delta \text{, then} \tag{33}$$
$$\sum_{w \in \Omega} P(w) \cdot \mathbf{I}[f_{C,\mathsf{G}^\mathsf{E}}(g_{\text{meta}}, w), w] < \sum_{w \in \Omega} P(w) \cdot \mathbf{I}[f_{P,\mathsf{G}^\mathsf{E}}(g, w), w]$$

Finally, since $f_{C,\mathsf{G}^\mathsf{E}}(g_{\text{meta}}, w) = f_{P,\mathsf{G}^\mathsf{E}}(g_{\text{acc-meta}}, w)$, (33) entails:

$$\text{There is a } \delta > \text{ such that, if } P(\textit{Guess Right}) > 1 - \delta \text{, then} \tag{34}$$
$$\sum_{w \in \Omega} P(w) \cdot \mathbf{I}[f_{P,\mathsf{G}^\mathsf{E}}(g_{\text{acc-meta}}, w), w] < \sum_{w \in \Omega} P(w) \cdot \mathbf{I}[f_{P,\mathsf{G}^\mathsf{E}}(g, w), w]$$

This completes the proof of Theorem 2.[25]

Theorem 2 says that, for a wide class of fallible subjects, and any evidential updating rule g that disagrees with Accurate Metaconditionalization, if the subject is sufficiently confident that she will correctly identify her evidence, then she will expect adopting Accurate Metaconditionalization to have strictly lower inaccuracy than adopting g.[26] Accurate Metaconditionalization, however, is not Bayesian Conditionalization. Importantly, this is so *whether or not evidence externalism is true*. Bayesian Metaconditionalization says that you should respond to your evidence in w by conditioning on your evidence in w, $\mathsf{E}(w)$. Accurate Metaconditionalization says that you should respond to your evidence in w by conditioning on the proposition *Guess Right and* $[\mathsf{E} = \mathsf{E}(w)]$. Since $\mathsf{E}(w)$

[25]Throughout this paper I have assumed that, for any evidential updating rule g, $f_{C,\mathsf{G}^\mathsf{E}}(g, w) = g(C, \mathsf{G}^\mathsf{E}(w))$. This assumption is not essential. Let *Follow Meta* be the proposition that if you adopted Metaconditionalization, you would follow Metaconditionalization. Let *Follow Cond* be the proposition that if you adopted Bayesian Conditionalization, you would follow Bayesian Conditionalization. Let *Follow Both* be the conjunction of *Follow Meta* and *Follow Cond*. Let C be your prior. Suppose that $C(\mathsf{E} = \mathsf{E}(w) | \textit{Follow Both}) > 0$ for all $w \in \Omega$. Then if you are sufficiently confident in *Follow Both*, you will expect adopting the following rule to be more accurate than adopting Bayesian Conditionalization.

> **Infallible Metaconditionalization**.
> Where C is any prior such that $C(\mathsf{E} = \mathsf{E}(w) | \textit{Follow Both}) > 0$ for all $w \in \Omega$:
> $g_{\text{i-meta}}(C, \mathsf{E}(w)) = C(\cdot | \textit{Follow Both} \wedge \mathsf{E} = \mathsf{E}(w))$

[26]In fact, the subject doesn't have to be *that* confident that she will correctly identify her evidence. In Schultheis (ms), I give a model of the unmarked clock in which anything over 50% confident will do.

need not entail *Guess Right*, internalists and externalists will agree that Accurate Metaconditionalization is not Bayesian Conditionalization.

Thus, Theorem 2 shows that the Bayesian Dilemma is a genuine dilemma, and that it is a dilemma for everyone. Everyone—internalists and externalists alike—must admit that, for a wide class of fallible subjects, if the subject is sufficiently confident that *Guess Right* is true, then she will expect Bayesian Metaconditionalization to be more inaccurate than Accurate Metaconditionalization. Everyone must choose between Accuracy-First Updating and the claim that Bayesian Conditionalization is always the rational update rule.[27]

5.2 Certainty Internalism

As I mentioned, I expect some will say that Theorem 2 nevertheless poses an additional, distinctive threat to externalism. They will say that if Accurate Metaconditionalization is the rational update rule for you, then it follows that, whenever your evidence is $\mathsf{E}(w)$, it is rational for you to be certain *that* your evidence is $\mathsf{E}(w)$. Let *Certainty Internalism* be the view that it is always rational to be certain of what your evidence is.

> **Certainty Internalism**
> If your evidence is $\mathsf{E}(w)$, then it is rational to be certain that your evidence is $\mathsf{E}(w)$.

I have argued that Accurate Metaconditionalization is the rational update procedure for a wide class of fallible subjects. If Accurate Metaconditionalization is rationally required for a wide class of fallible subjects, Certainty Internalism will also be true of those subjects.

But contemporary externalists tend to reject Certainty Internalism. Evidence externalism says that sometimes your evidence $\mathsf{E}(w)$ does not entail that your evidence is $\mathsf{E}(w)$. It would be natural to think, then, that the evidence externalist

[27]It is worth taking a moment to see how this result interacts with considerations of *availability* that are often discussed in the context of Schoenfield's result. We said that many theorists (implictly) take the accuracy-first thesis to be a thesis about which rule to follow. On this understanding, the thesis says, roughly, that we're rationally required to follow an updating rule that is such that (1) following that rule is an available option and (2) following that rule minimizes expected inaccuracy among the available options. In footnote 20 I said that the externalist should deny that Metaconditionalization is (always) an available option. My result does not assume that following Accurate Metaconditionalization (or Metaconditionalization for that matter) is an available option; I assume only that *adopting* Accurate Metaconditionalization is an available option. I see no reason principled reasons for denying that this is so. The externalist says that I cannot make it the case that I am always certain of the true answer to the question of what my evidence is. They do not deny that I can *try* or *plan* to be certain of true answer to the question of what my evidence.

must also say that sometimes, it's not rational to be *certain* that your evidence is $\mathsf{E}(w)$. Return to the case in which you're feeling cold, but your mechanisms specialized for detecting feelings of coldness misfire, telling you that you're not feeling cold. We've said that since you are feeling cold, it is not part of your evidence that you're not feeling cold. But you have no reason to believe that anything is amiss—no reason to believe it's not part of your evidence that you're not feeling cold. Surely, it's natural to say, you should not be certain that it is not part of your evidence that you're not feeling cold—by hypothesis, you have every reason to believe you're feeling cold, and hence every reason to believe it *is* part of your evidence that you're feeling cold.

Tempting as this sounds, I don't think the externalist is compelled to reject Certainty Internalism. The core commitment of externalism, as I have characterized the view, is that you do not always have access to what your evidence is; your evidence does not always tell you what your evidence is. (I take this to follow from a much more general commitment to a kind of *anti-Cartesianism*, according to which there is *no* special domain of facts—be they mental facts, normative facts, or any other kind of fact—to which we have guaranteed and perfect access.) It does *not* follow from this core commitment alone that Certainty Internalism is false. To derive the conclusion that Certainty Internalism is false, we also have to assume that it isn't rational to be certain of propositions that are not entailed by your evidence. But that assumption is not a core commitment of epistemic externalism, and as we've seen, it's not consistent with an accuracy-first approach to updating. The accuracy-first externalist can make her peace with Certainty Internalism.

6 Conclusion

It's been said that accuracy-first epistemology poses a special threat to externalism. Schoenfield (2017) shows that the rule that maximizes expected accuracy is Metaconditionalization. But if externalism is true, the argument goes, Metaconditionalization is not Bayesian Conditionalization. Thus, externalists face a dilemma, which I have called the *Bayesian Dilemma*: Either deny that Bayesian Conditionalization is required or else deny that the rational update rule is the rule that maximizes expected accuracy.

I am not convinced by these arguments. Schoenfield's result shows that *following* Metaconditionalization has greater expected accuracy than *following* Bayesian Conditionalization. It does not follow that *adopting* Metaconditionalization has greater expected accuracy than adopting Bayesian Conditionalization. That would follow only if we also said that if you adopted Metaconditionalization, you would follow Metaconditionalization. But the externalist has every reason to deny that this is always so.

I have argued that the Bayesian Dilemma is nevertheless a genuine dilemma. I presented a new argument that does not make any assumptions that the externalist must reject. This argument shows that, for a wide class of fallible subjects, if the subject is sufficiently confident that she will correctly identify her evidence, then Accurate Metaconditionalization will have greater expected accuracy than Conditionalization—and importantly, this is so for internalists and externalists alike. The Bayesian Dilemma is a genuine dilemma, and it is a dilemma for everyone.

7 Appendix

Lemma 1
Let E be any evidence function, G^{E} any guess function, and C any prior such that $C(\textit{Guess Right}) = 1$ and $C(\mathsf{E} = \mathsf{E}(w)) > 0$ for all $w \in \Omega$. Let P be any (possibly fallible) prior, and let g be any evidential updating rule such that $g(P, \mathsf{E}(w)) \neq g_{\text{meta}}(C, \mathsf{E}(w))$ for some $w \in \Omega$. Then:

$$\mathbb{E}_C[\mathbf{I}(f_{C,\mathsf{G}^{\mathsf{E}}}(g_{\text{meta}}))] < \mathbb{E}_C[\mathbf{I}(f_{P,\mathsf{G}^{\mathsf{E}}}(g))]$$

Proof.

Let $E_1, \ldots, E_n$ be the propositions that the subject might learn. Let $\mathcal{E} = \{[\mathsf{E} = E_1], \ldots, [\mathsf{E} = E_n]\}$. We know that for some E_j, $g(P, E_j)) \neq g_{\text{meta}}(C, E_j)$. Since $C(\mathsf{E} = E_i) > 0$ for all E_i, it follows from the fact that $\mathbf{I}$ is strictly proper that:

(a) For any E_i:

$$\sum_{w \in [\mathsf{E}=E_i]} C(w|\mathsf{E} = E_i) \cdot \mathbf{I}(C(\cdot|\mathsf{E} = E_i), w) \leq \sum_{w \in [\mathsf{E}=E_i]} C(w|\mathsf{E} = E_i) \cdot \mathbf{I}(g(P, E_i), w)$$

(b) For some E_j:

$$\sum_{w \in [\mathsf{E}=E_j]} C(w|\mathsf{E} = E_j) \cdot \mathbf{I}(C(\cdot|\mathsf{E} = E_j), w) < \sum_{w \in [\mathsf{E}=E_j]} C(w|\mathsf{E} = E_j) \cdot \mathbf{I}(g(P, E_j), w)$$

If we multiply both sides of (a) by $C(\mathsf{E} = E_i)$ and both sides of (b) by $C(\mathsf{E} = E_j)$, then we get (c) and (d), respectively:

(c) For any E_i:

$$\sum_{w \in [\mathsf{E}=E_i]} C(w) \cdot \mathbf{I}(C(\cdot|\mathsf{E} = E_i), w) \leq \sum_{w \in [\mathsf{E}=E_i]} C(w) \cdot \mathbf{I}(g(P, E_i), w)$$

(d) For some E_j:

$$\sum_{w \in [\mathsf{E}=E_j]} C(w) \cdot \mathbf{I}(C(\cdot|\mathsf{E} = E_j), w) < \sum_{w \in [\mathsf{E}=E_j]} C(w) \cdot \mathbf{I}(g(P, E_j), w)$$

We then sum over the propositions in the partition $\mathcal{E}$. This gives us (e):

(e) $$\sum_{[\mathsf{E}=E_i] \in \mathcal{E}} \sum_{w \in [\mathsf{E}=E_i]} C(w) \cdot \mathbf{I}(C(\cdot|\mathsf{E} = E_i), w) < \sum_{[\mathsf{E}=E_i] \in \mathcal{E}} \sum_{w \in [\mathsf{E}=E_i]} C(w) \cdot \mathbf{I}(g(P, E_i), w)$$

If $w \in [\mathsf{E} = E_i]$, then $E_i = \mathsf{E}(w)$. So (e) entails:

(f) $$\sum_{[\mathsf{E}=E_i] \in \mathcal{E}} \sum_{w \in [\mathsf{E}=E_i]} C(w) \cdot \mathbf{I}(C(\cdot|\mathsf{E} = \mathsf{E}(w)), w) < \sum_{[\mathsf{E}=E_i] \in \mathcal{E}} \sum_{w \in [\mathsf{E}=E_i]} C(w) \cdot \mathbf{I}(g(P, \mathsf{E}(w)), w)$$

Since $\mathcal{E}$ is a partition of Ω we can rewrite (f) as (g):

(g) $\sum_{w \in \Omega} C(w) \cdot \mathbf{I}(C(\cdot|\mathsf{E} = \mathsf{E}(w)), w) < \sum_{w \in \Omega} C(w) \cdot \mathbf{I}(g(P, \mathsf{E}(w)), w)$

By the definition of Metaconditionalization (g) entails:

(h) $\sum_{w \in \Omega} C(w) \cdot \mathbf{I}(g_{\text{meta}}(C, \mathsf{E}(w)), w) < \sum_{w \in \Omega} C(w) \cdot \mathbf{I}(g(P, \mathsf{E}(w)), w)$

Since $C(\textit{Guess Right}) = 1$, we know that if $C(w) > 0$, $\mathsf{E}(w) = \mathsf{G}^{\mathsf{E}}(w)$. So (h) entails:

(i) $\sum_{w \in \Omega} C(w) \cdot \mathbf{I}(g_{\text{meta}}(C, \mathsf{G}^{\mathsf{E}}(w)), w) < \sum_{w \in \Omega} C(w) \cdot \mathbf{I}(g(P, \mathsf{G}^{\mathsf{E}}(w)), w)$

We know that $g_{\text{meta}}(C, \mathsf{G}^{\mathsf{E}}(w)) = f_{C,\mathsf{G}^{\mathsf{E}}}(g_{\text{meta}}, w)$. So:

(j) $\sum_{w \in \Omega} C(w) \cdot \mathbf{I}(f_{C,\mathsf{G}^{\mathsf{E}}}(g_{\text{meta}}, w) < \sum_{w \in \Omega} C(w) \cdot \mathbf{I}(g(P, \mathsf{G}^{\mathsf{E}}(w)), w)$

And since g is an evidential update rule, we know that $g(P, \mathsf{G}^{\mathsf{E}}(w)) = f_{P,\mathsf{G}^{\mathsf{E}}}(g, w)$. So (j) entails:

(k) $\sum_{w \in \Omega} C(w) \cdot \mathbf{I}(f_{C,\mathsf{G}^{\mathsf{E}}}(g_{\text{meta}}, w) < \sum_{w \in \Omega} C(w) \cdot \mathbf{I}(f_{P,\mathsf{G}^{\mathsf{E}}}(g, w), w)$

This completes the proof of Lemma 1.

8 References

Bacon, Andrew. (forthcoming). "Actual Value in Decision Theory,"*Analysis*.

Briggs, R. A. & Pettigrew, Richard (2020). "An Accuracy-Dominance Argument for Conditionalization," *Nous*, 54 (1): 162-181.

Bronfman, Aaron (2014). "Conditionalization and not Knowing that one Knows," *Erkenntnis*, 79(4): 871-92.

Campbell-Moore, Catrin & Levinstein, Benjamin A. (2021). "Strict propriety is weak." *Analysis,* 81 (1): 8-13.

Carr, Jennifer (2019). "A modesty proposal," *Synthese*, 198 (4): 3581-3601.

Das, Nilanjan (2019). "Accuracy and Ur-Prior Conditionalization," *Review of Symbolic Logic*, 12 (1): 62-96.

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

Gallow, Dmitri (2019). "Updating for Externalists," *Noûs*, 55 (3): 487-516.

Greaves, Hilary & Wallace, David (2006)."Justifying Conditionalisation: conditionalisation maximizes expected epistemic utility," *Mind* 115 (459): 607-632.

Hedden, Brian (2012). "Options and the Subjective Ought," *Philosophical Studies*, 158 (2): 343-360

Jeffrey, Richard (1965). *The Logic of Decision*. Chicago: University of Chicago Press.

Isaacs, Yoaav & Russell, Jeffrey Sanford Russell. "Updating Without Evidence," *Nous*, forthcoming.

Jeffrey, Richard (1992). "Preferences among Preferences," in *Probability and the Art of Judgment*, Cambridge: Cambridge University Press.

Joyce, James (1998). "A Non-pragmatic Vindication of Probabilism," *Philosophy of Science*, 65 (4): 575-603.

Koon, Justis (2020). "Options Must be External," *Philosophical Studies*, 177 (5): 1175-1189.

Lewis, David (1981). "Causal Decision Theory," *Australian Journal of Philosophy*, 59 (1): 5-30.

McDowell, John (1982). "Criteria, Defeasibility, and Knowledge," *Proceedings of the British Academy*, 68: 455-479.

McDowell, John (2011). *Perception as a Capacity for Knowledge*. Milwaukee: Marquette University Press,

Pettigrew, Richard (2012). "Accuracy, Chance, and the Principal Principle," *Philosophical Review*, 121 (2): 241-275.

Pettigrew, Richard (2013). "Accuracy and Evidence," *Dialectica*, 67 (4): 579-596.

Pettigrew, Richard (2016). "Accuracy, Risk, and the Principle of Indifference," *Philosophy and Phenomenological Research*, 92 (1): 35-59

Pettigrew, Richard (2016). *Accuracy and the Laws of Credence*. Oxford: Oxford University Press.

Schoenfield, Miriam (2015). "Bridging Rationality and Accuracy," *Journal of Philosophy*, 112 (12): 633-657

Schoenfield, Miriam (2017). "Conditionalization does not Maximize Expected Accuracy," *Mind*, 126 (504): 1155-1187.

Salow, Bernhard (2019). "Elusive Externalism," *Mind*, 128 (510): 397-427.

Stalnaker, Robert. (1968). 'A Theory of Conditionals.' In Nicholas Rescher (ed.) *Studies in Logical Theory* (American Philosophical Quarterly Monographs 2), Oxford: Blackwell. pp. 98-112.

Weatherson, Brian (2011). "Stalnaker on Sleeping Beauty," *Philosophical Studies*, 155: 445-456.

Williamson, Timothy. (2000). *Knowledge and its Limits*. Oxford: Oxford University Press.

Zendajas Medina, Pablo (2022). "Just as Planned: Bayesianism, Externalism, and Plan Coherence," *Philosophical Imprint*, forthcoming.