Philosophy of Science (2025), **92**, 324–343
doi:10.1017/psa.2024.36

ARTICLE

Accurate Updating

Ginger Schultheis

University of Chicago Division of the Humanities, Chicago, IL, USA
Email: vks@uchicago.edu

(Received 30 January 2023; revised 31 March 2023; accepted 13 August 2024)

Abstract

A number of authors have observed that epistemic externalists seem to face a dilemma: Either deny that Conditionalization is the rational update rule, thereby rejecting traditional Bayesian epistemology, or deny that the rational update rule maximizes expected accuracy, thereby rejecting accuracy-first epistemology. Call this the Bayesian Dilemma. I'm not convinced by this argument. Once we make the premises explicit, we see that it relies on assumptions the externalist rejects. In this paper, I argue that the Bayesian Dilemma is nevertheless a genuine dilemma. My argument does not make any assumptions that the externalist rejects.

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 credences 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.[1] 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*.[2]

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.

[1] Joyce (1998).

[2] See Greaves and Wallace (2006) and Easwaran (2013). 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).

Schoenfield (2017) has shown that following the update rule *Metaconditionalization* maximizes expected accuracy.[3] However, as she and many other authors note, if externalism is true, Metaconditionalization is not Bayesian Conditionalization. Therefore, the externalist seems to face a dilemma: Either deny that 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*.[4]

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. 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 you would follow Metaconditionalization if you adopted Metaconditionalization, then you'll expect adopting a rule I'll call *Accurate Metaconditionalization* to be more accurate than adopting Bayesian Conditionalization.

I'll start in section 2 by introducing an accuracy-based framework for evaluating updating rules in terms of what I will call *actual inaccuracy*. In section 3, I'll introduce externalism. In section 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 section 5, I present the continuity argument showing that the Bayesian Dilemma is nevertheless a genuine dilemma. Section 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'll present my own preferred way. 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 to work with measures of inaccuracy rather than measures of accuracy. An *inaccuracy measure* $\mathbf{I}$ is a function that takes a world w from a set of worlds Ω, and a probability function C 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

[3] The name "Metaconditionalization" is due to Das (2019). I believe that this rule was first introduced and defended by Matthias Hild; see Hild (1998a,b). Hild calls the rule "Auto-Epistemic Conditionalization."

[4] For recent work on the relationship between accuracy-first epistemology and externalism, see Bronfman (2014), Schoenfield (2017), Das (2019), Gallow (2021), and Zendejas Medina (2023). Note that not all of these authors argue for the Bayesian Dilemma as I have presented it. For example, Zendejas Medina argues that the Bayesian Dilemma is not a genuine dilemma; in particular, he claims that the dilemma only arises if we accept a certain *plan coherence* principle, and he argues that we should reject this principle. Das (2019) focuses on the relationship between externalism and accuracy-first arguments for *Ur-Prior Conditionalization* (instead of the rule of Conditionalization).

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 three assumptions about inaccuracy measures. Though these assumptions are not incontrovertible, they are standard in the accuracy-first literature, and I will not say much to justify them.[5]

The first assumption is:

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*—roughly, that rational agents should be doing best, by their own lights, in their pursuit of accuracy.

The second assumption is *Additivity*, which says, roughly, that the total inaccuracy score of a credence function at a world is the sum of the inaccuracy scores of each of its individual credences. More precisely:

Additivity
For any $H \in \mathcal{P}(\Omega)$, there is a *local inaccuracy measure* i^H that takes a world $w \in \Omega$ and a credence $C(H)$ in the proposition H to a real number such that:

$$\mathbf{I}(C, w) = \sum_{H \in \mathcal{P}(\Omega)} i^H_w(C(H))$$

The third assumption is a continuity assumption for local inaccuracy measures. Specifically:

Continuity
$i^H_w(x)$ is a continuous function of x.

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 $\mathsf{E}(w)$:

[5] See, among others, Joyce (1998) and Pettigrew (2016) for defenses of Additivity and Continuity. See Joyce (1998) and Campbell-Moore and Levinstein (2021) for defenses of Strict Propriety.

https://doi.org/10.1017/psa.2024.36 Published online by Cambridge University Press

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

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.[6] In the next two sections of the paper, we will be talking about two updating rules. The first is Bayesian Conditionalization.

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* $E(w)$.

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 doing those doing things. Intending or planning 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.[7]

To see how this might happen, consider Williamson's well-known case of the unmarked clock.[8] 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

[6] 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, 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.

[7] My distinction between adopting a plan and following a plan is similar to Schoenfield's (2015) 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*. See also Isaacs and Russell (2023).

[8] Williamson (2000).

https://doi.org/10.1017/psa.2024.36 Published online by Cambridge University Press

clock in fact reads 4:05. Your 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.[9] 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 1 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 we do not want to say that we are rationally required to follow 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 (i) following that rule is an available option, and (ii) following that rule minimizes expected inaccuracy among the available options.[10] 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, if you tried to perform the act, you would.[11] But on this understanding, even following Bayesian Conditionalization is not always an available option, according to the externalist. 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 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 that 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

[9] This analysis of the case of the unmarked clock is due to Gallow (2021).

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

[11] For defenses of the view that the scope of our options is limited to the scope of our abilities, see Jeffrey (1965, 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.”

https://doi.org/10.1017/psa.2024.36 Published online by Cambridge University Press

adopt an available 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 the example of the unmarked clock suggests that cases like this are commonplace. We should take them into account. In light of this, I suggest that we understand 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*.[12] 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.[13] 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—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 simplfiying assumptions about how we are disposed to change our credal states if we adopt Bayesian Conditionalization or Metaconditionalization.

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

(I) If you adopted Bayesian Conditionalization, 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 (I) is false—if you adopted Bayesian Conditionalization you would condition on the wrong proposition. I will assume that these are the only two cases. Either you guess correctly and condition on the right proposition, or else you guess incorrectly and condition on the wrong proposition.[15]

[12] This term comes from Andrew Bacon's notion of *actual value*; see Bacon (2022).

[13] Note that when I talk about "worlds" I am talking about *big worlds*—maximally specific worlds that settle answers to all questions, including questions about what your evidence is and what credence function you adopt.

[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 $h(\text{Adopt} - g, w)$.

[15] In the main text I am assuming that when you adopt Bayesian Conditionalization, you follow a three-step process: (i) you receive some evidence, (ii) you guess what your evidence is, and (iii) you

https://doi.org/10.1017/psa.2024.36 Published online by Cambridge University Press

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.[16] Then, where $f_{C,\mathsf{G}^{\mathsf{E}}}$ is the credal selection function for any subject with guess function G^{E} and prior C:[17]

$$f_{C,\mathsf{G}^{\mathsf{E}}}\left(g_{\text{cond}}, w\right) = g_{\text{cond}}\left(C, \mathsf{G}^{\mathsf{E}}(w)\right), \tag{3}$$

$$f_{C,\mathsf{G}^{\mathsf{E}}}\left(g_{\text{meta}}, w\right) = g_{\text{meta}}\left(C, \mathsf{G}^{\mathsf{E}}(w)\right). \tag{4}$$

Equation (3) says that the credence function you would have if you adopted Bayesian Conditionalization in a world w, given that you have prior C and guess function G^{E}, is the result of conditioning your prior C on $\mathsf{G}^{\mathsf{E}}(w)$, your guess about what your evidence is in w. Likewise, (4) says that the credence function you would have if you adopted Metaconditionalization in a world w, given that you have prior C and guess function G^{E}, is the result of conditioning your prior C on the proposition *that your evidence is* $G^E(w)$, your guess about what your evidence is in w.

We will now use credal selection functions to define the *actual accuracy* of adopting an evidential updating rule. Let g be any evidential updating rule. Let G^{E} be any guess function. Let C be any prior. We define $V_{C,\mathsf{G}^{\mathsf{E}}}(g, w)$, the actual inaccuracy, in w, of adopting rule g given prior C and guess function G^{E}, as follows:

Actual Inaccuracy

$$V_{C,\mathsf{G}^{\mathsf{E}}}(g, w) = \mathbf{I}\left[f_{C,\mathsf{G}^{\mathsf{E}}}(g, w), w\right].$$

The actual inaccuracy, in w, of adopting the updating rule g given that you have guess function G^{E} and prior C is the inaccuracy, in w, of the credence function you would have if you adopted rule g in w, given that C is your prior and G^{E} is your guess function.[18]

condition your prior on your guess. But it is far from obvious that adopting an updating rule always involves the intermediate step (ii). As an anonymous referee points out, it may be that you simply *respond* to your evidence without forming any (explicit or implicit) beliefs *about* what your evidence is. In the end, I want to agree with this. I do not think that the externalist *has* to think of adopting Conditionalization as involving my intermediate step (ii). What I *do* think is that the kinds of motivations that lead us to accept externalism should also lead us to believe that, at least sometimes, you will adopt Conditionalization as your update rule yet fail to follow Conditionalization because you condition your prior on the wrong proposition. If we accept that your evidence can come apart from the proposition that you condition on in this way, we can think of G^{E} as representing the proposition that you condition on (in the case of Conditionalization). For simplicity, I will continue to talk about guesses about your evidence in the main text, but it is important to remember that the formalism does not have to be interpreted in this way.

[16] Isaacs and Russell (2023) 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 and Russell, but I do not have the space to address them here.

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

[18] 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.

https://doi.org/10.1017/psa.2024.36 Published online by Cambridge University Press

Assuming (3), the actual inaccuracy of adopting Bayesian Conditionalization in a world w for a subject with prior C and guess function G^{E} is

$$\mathbf{I}\left[f_{C,\mathsf{G}^{\mathsf{E}}}\left(g_{\text{cond}}, w\right), w\right] = \mathbf{I}\left[g_{\text{cond}}\left(C, \mathsf{G}^{\mathsf{E}}(w)\right), w\right]. \quad (5)$$

Assuming (4), the actual inaccuracy of adopting Metaconditionalization in a world w for a subject with prior C and guess function G^{E} is

$$\mathbf{I}\left[f_{C,\mathsf{G}^{\mathsf{E}}}\left(g_{\text{meta}}, w\right), w\right] = \mathbf{I}\left[g_{\text{meta}}\left(C, \mathsf{G}^{\mathsf{E}}(w)\right), w\right]. \quad (6)$$

The *expected actual inaccuracy* of adopting Bayesian Conditionalization and of adopting Metaconditionalization are defined in (7) and (8), respectively:

$$\sum_{w \in \Omega} C(w) \cdot \mathbf{I}\left[f_{C,\mathsf{G}^{\mathsf{E}}}\left(g_{\text{cond}}, w\right), w\right] = \sum_{w \in \Omega} C(w) \cdot \mathbf{I}\left[g_{\text{cond}}\left(C, \mathsf{G}^{\mathsf{E}}(w)\right), w\right], \quad (7)$$

$$\sum_{w \in \Omega} C(w) \cdot \mathbf{I}\left[f_{C,\mathsf{G}^{\mathsf{E}}}\left(g_{\text{meta}}, w\right), w\right] = \sum_{w \in \Omega} C(w) \cdot \mathbf{I}\left[g_{\text{meta}}\left(C, \mathsf{G}^{\mathsf{E}}(w)\right), w\right]. \quad (8)$$

Returning 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:

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 if, 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 $\mathsf{E}(w)$.

https://doi.org/10.1017/psa.2024.36 Published online by Cambridge University Press

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 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 my 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.[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. Consider any updating rule g and any prior C such that $g(C, E(w)) \neq g_{meta}(C, E(w))$ for some w such that $C(w) > 0$. Then*

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

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

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

https://doi.org/10.1017/psa.2024.36 Published online by Cambridge University Press

open worlds where g and Metaconditionalization disagree. Then, Theorem 1 says, that 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, Baysian 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 always the same proposition as $[\mathsf{E} = \mathsf{E}(w)]$. In particular, sometimes $\mathsf{E}(w)$ will not entail the proposition $[\mathsf{E} = \mathsf{E}(w)]$, and when this happens, Metaconditionalization and Bayesian Conditionalization will disagree.

Let E be any learning situation in which $[\mathsf{E} = \mathsf{E}(w)] \neq \mathsf{E}(w)$ for some world w. Consider any subject who leaves open some such worlds. Then Theorem 1 entails that the subject will expect the recommendation of Metaconditionalization to be less inaccurate than the recommendation of Bayesian Conditionalization in learning situation E. Formally:

$$\sum_{w\in\Omega} C(w)\cdot \mathbf{I}[g_{\text{meta}}(C,\mathsf{E}(w)] < \sum_{w\in\Omega} C(w)\cdot \mathbf{I}[g_{\text{cond}}(C,\mathsf{E}(w)]. \tag{9}$$

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

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

$$\textit{Guess Right} = \left\{w\in\Omega : \mathsf{G}^{\mathsf{E}}(w) = \mathsf{E}(w)\right\}. \tag{10}$$

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

$$\textit{Guess Wrong} = \left\{w\in\Omega : \mathsf{G}^{\mathsf{E}}(w) \neq \mathsf{E}(w)\right\}. \tag{11}$$

Say that a subject with guess function G^{E} is *infallible* in learning situation E if, for any $w \in \Omega$, *Guess Right* is true in w. If we assume that our subject is infallible in learning situation E, then for all $w \in \Omega$,

$$f_{C,\mathsf{G}^{\mathsf{E}}}\left(g_{\text{cond}}, w\right) = g_{\text{cond}}(C,\mathsf{E}(w)), \tag{12}$$

$$f_{C,\mathsf{G}^{\mathsf{E}}}\left(g_{\text{meta}}, w\right) = g_{\text{meta}}(C,\mathsf{E}(w)). \tag{13}$$

If (12) and (13) are true, then Theorem 1 entails that the subject expects adopting Metaconditionalization to be less inaccurate than adopting Bayesian Conditionalization. Formally:

https://doi.org/10.1017/psa.2024.36 Published online by Cambridge University Press

$$\sum_{w \in \Omega} C(w) \cdot \mathbf{I}\Big[f_{C,\mathsf{G}^{\mathsf{E}}}\big(g_{\text{meta}}, w\big), w\Big] < \sum_{w \in \Omega} C(w) \cdot \mathbf{I}\Big[f_{C,\mathsf{G}^{\mathsf{E}}}\big(g_{\text{cond}}, w\big), w\Big]. \tag{14}$$

But of course the externalist will insist that creatures like us are not infallible. Remember, the externalist says my beliefs about what evidence I have are not perfectly sensitive to the facts about what evidence I have. Return to the case of the unmarked clock. In fact my evidence is that the time is between 4:04 and 4:06. But my mechanisms specialized for detecting what evidence I have misfire, and so I mistakenly think that my evidence is some other proposition—that the time is between 4:05 and 4:07. Importantly, the externalist maintains that no amount of careful attention to my evidence will insure me against error. For the externalist, even ideally rational, maximally attentive agents are not always certain of the true answer to the question of what their evidence is. That is just to say that even ideally rational, maximally attentive agents are not always such that, if they adopted Metaconditionalization, they would follow Metaconditionalization.

In short, (13) is often false for agents like us—agents with fallible information-gathering mechanisms. But without (13), we can't derive (14) from (9). We can't conclude that, for fallible agents like us, adopting Metaconditionalization has lower expected actual inaccuracy than adopting Bayesian Conditionalization.

Let me summarize. If evidence externalism is true, then Theorem 1 tells us that, under certain conditions, we will expect following Metaconditionalization to be less inaccurate than following any other evidential updating rule. It doesn't follow, however, that we expect *adopting* Metaconditionalization to be less inaccurate than adopting any other rule.[20] In particular, it doesn't follow that we expect adopting Metaconditionalization to be less inaccurate than adopting Bayesian Conditionalization. That would follow only if we knew that we're infallible, but we cannot, on pain of begging the question against the externalist, simply assume that this is so. So we have not shown that if evidence externalism is true, then we must choose between the rule that maximizes expected accuracy and Bayesian Conditionalization.[21]

[20] Steel (2018) makes this same point in a different context. He observes that the Greaves and Wallace accuracy argument for Bayesian Conditionalization at best shows that Bayesian Conditionalization is the optimal rule to *follow*; it does not show that Bayesian Conditionalization is the optimal rule to *try to follow*.

[21] 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 section 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 (i) the claim that we're required to *follow* Bayesian Conditionalization and (ii) 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, for the reasons I discuss in the main text, the externalist should deny that we are always able to follow Metaconditionalization.

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 that you will correctly identify your evidence, then you will expect a rule that I call *Accurate Metaconditionalization* to have less expected inaccuracy than adopting Bayesian Conditionalization. In section 5.1 I'll begin by saying what Accurate Metaconditionalization is, and then I'll present the continuity argument. In section 5.2 I will consider whether other rules are immune to the continuity argument.

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. (Remember, $Guess\ \ Right = \{w \in \Omega : \mathsf{G}^{\mathsf{E}}(w) = \mathsf{E}(w)\}$.) More precisely:

Accurate Metaconditionalization

Where C is any prior such that $C(\mathsf{E} = \mathsf{E}(w)|Guess\ \ Right) > 0$ for all $w \in \Omega$,

$$g_{\text{acc-meta}}(C, \mathsf{E}(w)) = C(\cdot|Guess\ \ Right \wedge \mathsf{E} = \mathsf{E}(w)).$$

For simplicity, I will assume:

$$f_{C,\mathsf{G}^{\mathsf{E}}}\left(g_{\text{acc-meta}}, w\right) = g_{\text{acc-meta}}\left(C, \mathsf{G}^{\mathsf{E}}(w)\right) = C(\cdot|Guess\ \ Right \wedge \mathsf{E} = \mathsf{G}^{\mathsf{E}}(w)). \qquad (15)$$

Equation (15) 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 that she will correctly identify her evidence, then adopting Accurate Metaconditionalization will have lower expected actual inaccuracy than adopting Bayesian Conditionalization for her. Here is roughly how the argument will 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 will show that when $x = 1$, adopting Bayesian Conditionalization has greater expected actual inaccuracy than adopting Accurate Metaconditionalization. Since both functions are continuous, it follows there is some $\delta > 0$ such that if $x > 1 - \delta$, 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 kind of function, which I'll call a *probability extension function*. 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

https://doi.org/10.1017/psa.2024.36 Published online by Cambridge University Press

$\{Guess\ Right, Guess\ Wrong\}$ 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. Let Δ be the set of probability functions over $\mathcal{P}(\Omega)$. We define Δ_{Right} as

$$\Delta_{\text{Right}} = \{P_{\text{R}} : P_{\text{R}} \in \Delta \text{ and } P_{\text{R}}(Guess\ \ Right) = 1\}, \tag{16}$$

and we define Δ_{Wrong} in a similar way:

$$\Delta_{\text{Wrong}} = \{P_{\text{W}} : P_{W} \in \Delta \text{ and } P_{\text{W}}(Guess\ \ Wrng) = 1\}. \tag{17}$$

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

$$\lambda_{\langle P_{\text{R}}, P_{\text{W}}\rangle}(x)(\cdot) = P_{\text{R}}(\cdot)x + P_{\text{W}}(\cdot)(1 - x). \tag{18}$$

Each probability extension function is indexed to a pair $\langle P_{\text{R}}, P_{\text{W}}\rangle$. In what follows I will leave off the subscripts for the sake of readability.

We can 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}, and an evidential updating rule g. Each probability extension function λ determines a function that takes a credence x in *Guess Right* and returns the expectation, relative to $\lambda(x)$, of the actual inaccuracy of adopting rule g given guess function G^{E}. For example, consider

$$\sum_{w \in \Omega} \lambda(x)(w) \cdot \mathbf{I}\left[f_{\lambda(x),\mathsf{G}^{\mathsf{E}}}(g_{\text{meta}}, w), w\right] = \sum_{w \in \Omega} \lambda(x)(w) \cdot \mathbf{I}[g_{meta}(\lambda(x), \mathsf{G}^{\mathsf{E}}(w)), w]. \tag{19}$$

This is a function that takes a credence x in *Guess Right* and returns the expectation, relative to $\lambda(x)$, of the actual inaccuracy of adopting Metaconditionalization given guess function G^{E}. Similarly, we have

$$\sum_{w \in \Omega} \lambda(x)(w) \cdot \mathbf{I}\left[f_{\lambda(x),\mathsf{G}^{\mathsf{E}}}(g_{\text{cond}}, w), w\right] = \sum_{w \in \Omega} \lambda(x)(w) \cdot \mathbf{I}[g_{\text{cond}}(\lambda(x), \mathsf{G}^{\mathsf{E}}(w)), w]. \tag{20}$$

This is a function that takes a credence x in *Guess Right* and returns the expectation, relative to $\lambda(x)$, of the actual inaccuracy of adopting Bayesian Conditionalization given guess function G^{E}.

For any probability extension function λ, we define $\mathcal{C}_{\lambda}$ as follows:

$$\mathcal{C}_{\lambda} = \{C \in \Delta : C = \lambda(C(Guess\ Right))\}. \tag{21}$$

We're thinking of λ as a specification of the conditional credences of some hypothetical subject, conditional on each member of $\{Guess\ Right, Guess\ \ Wrong\}$ that the subject leaves open. We can then think of $\mathcal{C}_{\lambda}$ as the set of all probability functions that agree with λ with respect to those assignments of conditional credences.

Importantly, every probability function $C \in \Delta$ belongs to $\mathcal{C}_\lambda$ for some probability extension function λ.[22]

We will show that for any probability extension function λ satisfying certain constraints, and any probability function C in $\mathcal{C}_\lambda$, if $C(\textit{Guess Right})$ is sufficiently high, then the expected actual inaccuracy, relative to C, of adopting Accurate Metaconditionalization will be lower than the expected actual inaccuracy of adopting Bayesian Conditionalization. More precisely:

Theorem 2. *Let E be any learning situation, G^E any guess function, and λ any probability extension function such that:*

(i) $\lambda(1)(\mathsf{E}(w)) > 0$ for all $w \in \Omega$;
(ii) $\lambda(1)(\mathsf{E} = \mathsf{E}(w)) > 0$ for all $w \in \Omega$;
(iii) $g_{\text{meta}}(\lambda(1), \mathsf{E}(w)) \neq g_{\text{cond}}(\lambda(1), \mathsf{E}(w))$ for some $w \in \textit{Guess Right}$.

Then there's a $\delta_\lambda > 0$ such that, for all $C \in \mathcal{C}_\lambda$, if $C(\textit{Guess Right}) > 1 - \delta_\lambda$, then

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

The proof of Theorem 2 relies on a lemma.

Lemma 1 *Let E be any learning situation, G^E any guess function, and λ any probability extension function satisfying conditions (i) and (ii) in our statement of Theorem 2. Then*

(i) $\sum_{w \in \Omega} \lambda(x)(w) \cdot \mathbf{I}\Big[f_{\lambda(1),\mathsf{G}^\mathsf{E}}\big(g_{\text{meta}}, w\big), w\Big)]$
(ii) $\sum_{w \in \Omega} \lambda(x)(w) \cdot \mathbf{I}\Big[f_{\lambda(x),\mathsf{G}^\mathsf{E}}\big(g_{cond}, w\big), w\Big)]$

are both continuous at 1.

I leave the proof of Lemma 1 to the appendix.

Proof of Theorem 2. Consider any learning situation E, any guess function G^E, and any probability extension function λ satisfying (i), (ii), and (iii). It follows from Theorem 1 that

$$\sum_{w \in \Omega} \lambda(1)(w) \cdot \mathbf{I}\big[g_{\text{meta}}(\lambda(1), \mathsf{E}(w)), w\big] < \sum_{w \in \Omega} \lambda(1)(w) \cdot \mathbf{I}\big[g_{\text{cond}}(\lambda(1), \mathsf{E}(w)), w\big]. \quad (22)$$

This says that any subject whose prior is $\lambda(1)$ expects following Metaconditionalization in learning situation E to have lower expected inaccuracy than following Bayesian Conditionalization in learning situation E. Note that $\lambda(1)(\textit{Guess Right}) = 1$. This means that, for all $w \in \Omega$ such that $\lambda(1)(w) > 0$,

$$g_{\text{meta}}(\lambda(1), \mathsf{E}(w)) = f_{\lambda(1),\mathsf{G}^\mathsf{E}}\big(g_{\text{meta}}, w\big), \quad (23)$$

[22] If $C(\textit{Guess Right}) > 0$ and $C(\textit{Guess Wrong}) > 0$, then let $\lambda = \lambda_{\langle P_R, P_W \rangle}$ where $P_R(\cdot) = C(\cdot|\textit{Guess Right})$ and $P_W(\cdot) = C(\cdot|\textit{Guess Wrong})$. If $C(\textit{Guess Wrong}) = 1$, then let $\lambda = \lambda_{\langle P_R, P_W \rangle}$ where P_R is any probability function in Δ_{right}, and $P_W(\cdot) = C(\cdot)$. If $C(\textit{Guess Right}) = 1$, let $\lambda = \lambda_{\langle P_R, P_W \rangle}$ where P_W is any probability function in Δ_{wrong} and $P_R(\cdot) = C(\cdot)$.

$$g_{\text{cond}}(\lambda(1), \mathsf{E}(w)) = f_{\lambda(1),\mathsf{G}^{\mathsf{E}}}(g_{\text{cond}}, w). \tag{24}$$

Given (23) and (24), (22) entails

$$\sum_{w\in\Omega} \lambda(1)(w) \cdot \mathbf{I}\Big[f_{\lambda(1),\mathsf{G}^{\mathsf{E}}}(g_{\text{meta}}, w), w\Big] < \sum_{w\in\Omega} \lambda(1)(w) \cdot \mathbf{I}\Big[f_{\lambda(1),\mathsf{G}^{\mathsf{E}}}(g_{\text{cond}}, w), w\Big]. \tag{25}$$

This says that any subject whose prior is $\lambda(1)$ and whose guess function is G^{E} expects adopting Metaconditionalization in learning situation E to have lower expected inaccuracy than adopting Bayesian Conditionalization in learning situation E.

Equation (25) and Lemma 1 together entail that

$$\sum_{w\in\Omega} \lambda(x)(w) \cdot \mathbf{I}\Big[f_{\lambda(1),\mathsf{G}^{\mathsf{E}}}(g_{\text{meta}}, w), w\Big] < \sum_{w\in\Omega} \lambda(x)(w) \cdot \mathbf{I}\Big[f_{\lambda(x),\mathsf{G}^{\mathsf{E}}}(g_{\text{cond}}, w), w\Big]. \tag{26}$$

We know that for all $C \in \mathcal{C}_\lambda$, $C = \lambda\big(C\big(Guess\ \ Right\big)\big)$. Therefore, it follows from (26) that

There's a $\delta_\lambda > 0$s.t., for all $C \in \mathcal{C}_\lambda$, if $\mathrm{C}\big(Guess\ \ Right\big) > 1 - \delta_\lambda$, then

$$\sum_{w\in\Omega} C(w) \cdot \mathbf{I}\Big[f_{\lambda(1),\mathsf{G}^{\mathsf{E}}}(g_{\text{meta}}, w), w\Big] < \sum_{w\in\Omega} C(w) \cdot \mathbf{I}\Big[f_{C,\mathsf{G}^{\mathsf{E}}}(g_{\text{cond}}, w), w\Big]. \tag{27}$$

This says that for any subject whose prior probability functions is in $\mathcal{C}_\lambda$, if the subject is sufficiently confident in *Guess Right*, then she will expect adopting Metaconditionalization *with respect to* $\lambda(1)$ to have strictly lower actual inaccuracy than adopting Bayesian Conditionalization with respect to her own prior. Remember, we're assuming that

$$f_{C,\mathsf{G}^{\mathsf{E}}}(g_{\text{acc-meta}}, w) = g_{\text{acc-meta}}\big(C, \mathsf{G}^{\mathsf{E}}(w)\big) = C(\cdot|Guess\ \ Right \wedge \mathsf{E} = \mathsf{G}^{\mathsf{E}}(w)). \tag{28}$$

We are also assuming that

$$f_{C,\mathsf{G}^{\mathsf{E}}}(g_{\text{meta}}, w) = g_{\text{meta}}\big(C, \mathsf{G}^{\mathsf{E}}(w)\big) = C(\cdot|\mathsf{E} = \mathsf{G}^{\mathsf{E}}(w)). \tag{29}$$

It follows that

$$f_{C,\mathsf{G}^{\mathsf{E}}}(g_{\text{acc-meta}}, w) = f_{C(\cdot\,|\,Guess\ \ Right),\mathsf{G}^{\mathsf{E}}}(g_{\text{meta}}, w). \tag{30}$$

We know that, for all $C \in \mathcal{C}_\lambda$, if $C\big(Guess\ \ Right\big) > 0$ then

$$C(\cdot|Guess\ \ Right) = \lambda(1). \tag{31}$$

Equations (28) and (29) together entail that for all $C \in \mathcal{C}_\lambda$, if $C\big(Guess\ \ Right\big) > 0$, then

$$f_{C,\mathsf{G}^{\mathsf{E}}}(g_{\text{acc-meta}}, w) = f_{\lambda(1),\mathsf{G}^{\mathsf{E}}}(g_{\text{meta}}, w). \tag{32}$$

Given (30), (27) entails

There's a $\delta_\lambda > 0$ s.t., for all $C \in \mathcal{C}_\lambda$, if $\mathrm{C}\big(Guess\ \ Right\big) > 1 - \delta_\lambda$, then

$$\sum_{w\in\Omega} C(w) \cdot \mathbf{I}\Big[f_{C,\mathsf{G}^{\mathsf{E}}}(g_{\text{acc-meta}}, w), w\Big] < \sum_{w\in\Omega} C(w) \cdot \mathbf{I}\Big[f_{C,\mathsf{G}^{\mathsf{E}}}(g_{\text{cond}}, w), w\Big]. \tag{33}$$

This says that for any subject whose prior probability functions is in $\mathcal{C}_\lambda$ and whose guess function is G^{E}, if the subject is sufficiently confident in *Guess Right*, then she will expect adopting Accurate Metaconditionalization in learning situation E to have

https://doi.org/10.1017/psa.2024.36 Published online by Cambridge University Press

strictly lower actual inaccuracy than adopting Bayesian Conditionalization in learning situation E. This completes the proof of Theorem 2.

Let's take stock. In section 4, I presented Schoenfield's showing that *following* Metaconditionalization has greater expected actual accuracy than following Bayesian Conditionalization. But, I argued, we cannot conclude from this fact that *adopting* Metaconditionalization has greater expected actual accuracy than adopting Bayesian Conditionalization. That would follow only if we said that we're infallible in every learning situation, and we cannot, on pain of begging the question against the externalist, assume that this is so. In this section I have shown that we can do without the assumption of infallibility. Theorem 2 shows that for a wide class of fallible subjects and learning situations, if the subject is sufficiently confident that she will correctly identify her evidence in that learning situation, then adopting Accurate Metaconditionalization will have greater expected actual accuracy for her than adopting Bayesian Conditionalization.[23]

This is not good news for the project of reconciling accuracy-first externalism with Bayesian epistemology. The externalist who wishes to justify Bayesian Conditionalization on the basis of accuracy should hope to find a natural class of fallible agents for whom Bayesian Conditionalization is the most accurate updating procedure in expectation. We should be pessimistic about the prospects for this project on the basis of the results of this paper. Theorem 2 shows that adopting Accurate Metaconditionalization will have greater expected actual accuracy than adopting Conditionalization for some agents in any such class—so long as it includes agents who are sufficiently confident that they will correctly identify their evidence, and I can see no principled reason to exclude all such agents.

5.2 Guess conditionalization

Let me take a moment to address a concern about the significance of this result, and its relationship to other results in the literature. Those who have read Gallow (2021) or Isaacs and Russell (2023) might wonder: Haven't these authors already shown us how fallible agents should update their credences? Gallow (2021) argues that we can

[23] It is worth emphasizing that you don't have to be *that* confident that you will correctly identify your evidence. There are models of the unmarked clock in which anything over 50% will do. It is also 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 (i) following that rule is an available option, and (ii) following that rule minimizes expected inaccuracy among the available options. In footnote 4 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 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 the true answer to the question of what my evidence is. It is also worth noting that the results in this section do not depend on any assumptions about the structure of evidence. In particular, I have not assumed that evidence obeys introspection principles and I have not assumed that evidence is factive.

https://doi.org/10.1017/psa.2024.36 Published online by Cambridge University Press

use a version of a result due to Greaves and Wallace (2006) to show that a rule that I will call *Guess Conditionalization* is the best rule for fallible agents.[24]

Guess Conditionalization

$$g_{\mathsf{G-cond}}\big(C, \mathsf{G}^{\mathsf{E}}(w)\big) = C(\cdot|\mathsf{G}^{\mathsf{E}} = \mathsf{G}^{\mathsf{E}}(w)).$$

However, I believe that the argument that Gallow is alluding to requires certain assumptions about the nature of our fallibility that the externalist should reject. To see this, remember that our guess function G^{E} is a function that takes each world w to the subject's guess, in w, about what her evidence is. If we are interested in subjects who are trying to follow Guess Conditionalization, we need another guess function $\mathsf{G}^{\mathsf{G}^{\mathsf{E}}}$ that takes each world w to the subject's guess, in w, *about what her guess is in w*. Let us assume that

$$f_{C,\mathsf{G}^{\mathsf{G}^{\mathsf{E}}}}\big(g_{\mathsf{G-cond}}, w\big) = g_{\mathsf{G-cond}}\Big(C, \mathsf{G}^{\mathsf{G}^{\mathsf{E}}}(w)\Big). \tag{34}$$

This says that the credence function you would have if you adopted Guess Conditionalization in learning situation E is the result of conditioning your prior on your guess about what your guess is.[25] With this assumption in place, the Greaves and Wallace–style argument that Guess Conditionalization is the best rule for fallible agents requires us to assume that subjects are *guess-infallible*:

$$\text{For all worlds } \ w, \ \ \mathsf{G}^{\mathsf{G}^{\mathsf{E}}}(w) = \mathsf{G}^{\mathsf{E}}(w). \tag{35}$$

If we assume that our subject is guess-infallible then, for all $w \in \Omega$,

$$f_{C,\mathsf{G}^{\mathsf{G}^{\mathsf{E}}}}\big(g_{\mathsf{G-cond}}, w\big) = g_{\mathsf{G-cond}}\big(C, \mathsf{G}^{\mathsf{E}}(w)\big). \tag{36}$$

This says that if the subject adopts Guess Conditionalization, then she would follow Guess Conditionalization.

But the externalist should insist that creatures like us are not guess-infallible. According to the externalist, my beliefs about what I have guessed are not perfectly

[24] Note that Gallow himself is actually interested in a slightly different rule, which he calls *Update Conditionalization*. The differences between Update Conditionalization and Guess Conditionalization do not matter for my purposes.

[25] In the main text I assume (i) that you use your *first-order guesses* (your guesses about what your evidence is) when you try to follow Conditionalization, Metaconditionalization, or Accurate Conditionalization, and (ii) that you use your second-order guesses (your guesses about what your guesses are) when you try to follow Guess Conditionalization. One might question this assumption: Why couldn't I use my first-order guesses for Guess Conditionalization, too? I agree that my assumption that we always use our second-order guesses for Guess Conditionalization is not necessarily true, and I have made this assumption primarily to simplify the presentation of the argument in the main text. To avoid the argument that Guess Conditionalization is best, we do not have to assume that you *always* use your second-order guesses when you are trying to follow Guess Conditionalization. Rather, all we have to assume is that it is not always true that if you tried to follow Guess Conditionalization, then you would. For example, here is one way to say that this conditional is not always true without assuming that the agent *always* uses her second-order guesses when she tries to follow Guess Conditionalization. We could say that in some worlds where the agent's first-order guesses and second-order guesses come apart, she will try to follow Guess Conditionalization by using her first-order guesses, and in other worlds where her first-order guesses and her second-order guesses come apart, she will try to follow Guess Conditionalization by using her second-order guesses. Thanks to an anonymous referee.

sensitive to the facts about what I have guessed, and, importantly, no amount of careful attention to my guesses will insure me against error. Even ideally rational, maximally attentive agents are not always certain of the true answer to the question of what their guess is. That is just to say that even ideally rational, maximally attentive agents are not always such that, if they adopted Guess Conditionalization, they would follow Guess Conditionalization. In short, (34) is often false for agents like us—agents with fallible information-gathering mechanisms. But without (34), we can't use the Greaves and Wallace–style argument that Gallow is alluding to in order to show that adopting Guess Conditionalization has lower expected actual inaccuracy than adopting any other rule.

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, Metaconditionalization is not Bayesian Conditionalization. Thus, externalists seem to 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 this argument. Schoenfield's result shows that *following* Metaconditionalization has greater expected accuracy than following Bayesian Conditionalization. It doesn't 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 adopting Accurate Metaconditionalization will have greater expected accuracy for her than adopting Bayesian Conditionalization.

Acknowledgments. Thanks to David Boylan, Kevin Dorst, Matt Mandelkern, Alex Meehan, Dmitri Gallow, and Bernhard Salow 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.

References

Bacon, Andrew. 2022. "Actual Value in Decision Theory." *Analysis* 82 (4):617–29. https://doi.org/10.1093/analys/anac014

Briggs, R. A. and Richard Pettigrew. 2020. "An Accuracy-Dominance Argument for Conditionalization." *Nous* 54 (1):162–81. https://www.doi.org/10.1111/nous.12258

Bronfman, Aaron. 2014. "Conditionalization and Not Knowing That One Knows." *Erkenntnis* 79 (4):871–92. https://doi.org/10.1007/s10670-013-9570-0

Campbell-Moore, Catrin and Benjamin A. Levinstein. 2021. "Strict Propriety Is Weak." *Analysis* 81 (1):8–13. https://www.doi.org/10.1093/analys/anaa001

Das, Nilanjan. 2019. "Accuracy and Ur-Prior Conditionalization." *Review of Symbolic Logic* 12 (1):62–96. https://doi.org/10.1017/S1755020318000035

Easwaran, Kenny. 2013. "Expected Accuracy Supports Conditionalization—and Conglomerability and Reflection." *Philosophy of Science* 80 (1):119–142. https://doi.org/10.1086/668879
Gallow, Dmitri. 2021. "Updating for Externalists." *Noûs* 55 (3):487–516. https://doi.org/10.1111/nous.12307
Greaves, Hilary and David Wallace. 2006."Justifying Conditionalisation: Conditionalisation Maximizes Expected Epistemic Utility." *Mind* 115 (459):607–32. https://doi.org/10.1093/mind/f
Hedden, Brian. 2012. "Options and the Subjective Ought." *Philosophical Studies* 158 (2):343–60. https://doi.org/10.1007/s11098-012-9880-0
Hild, Matthias. 1998a. "Auto-Epistemology and Updating." *Philosophical Studies* 92 (3):321–61. https://doi.org/10.1023/A:1004229808144
Hild, Matthias. 1998b. "The Coherence Argument Against Conditionalization." *Synthese* 115 (2):229–58. https://doi.org/10.1023/A:1005082908147
Isaacs, Yoaav and Jeffrey Sanford Russell. 2023. "Updating Without Evidence." *Nous* 57 (3):576–99. https://doi.org/10.1111/nous.12426
Jeffrey, Richard. 1965. *The Logic of Decision*. Chicago IL: University of Chicago Press.
Jeffrey, Richard. 1992. "Preference Among Preferences." In *Probability and the Art of Judgment*, 154–69. 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):1175–89. https://doi.org/10.1007/s11098-019-01240-0
Lewis, David. 1981. "Causal Decision Theory." *Australian Journal of Philosophy* 59 (1):5–30. https://doi.org/10.1080/00048408112340011
McDowell, John. 1982. "Criteria, Defeasibility, and Knowledge." *Proceedings of the British Academy* 68:455–79.
McDowell, John. 2011. *Perception as a Capacity for Knowledge*. Milwaukee, WI: Marquette University Press.
Nielsen, Michael. 2021. Accuracy-Dominance and conditionalization. *Philosophical Studies* 178 (10):3217–3236.
Pettigrew, Richard. 2016. *Accuracy and the Laws of Credence*. Oxford: Oxford University Press.
Salow, Bernhard. 2019. "Elusive Externalism." *Mind* 128 (510):397–427. https://doi.org/10.1093/mind/fzx015
Schoenfield, Miriam. 2015. "Bridging Rationality and Accuracy." *Journal of Philosophy* 112 (12):633–57. https://doi.org/10.5840/jphil20151121242
Schoenfield, Miriam. 2017. "Conditionalization Does Not Maximize Expected Accuracy." *Mind* 126 (504):1155–87. https://doi.org/10.1093/mind/fzw027
Stalnaker, Robert. 1968. "A Theory of Conditionals." In *Studies in Logical Theory*, edited by Nicholas Rescher, 98–112. Oxford: Blackwell.
Steel, Robert. 2018. "Anticipating Failure and Avoiding It." *Philosophers' Imprint* 18 (14):1–28.
Weatherson, Brian. 2011. "Stalnaker on Sleeping Beauty." *Philosophical Studies* 155:445–56. https://doi.org/10.1007/s11098-010-9613-1
Williamson, Timothy. 2000. *Knowledge and its Limits*. Oxford: Oxford University Press.
Zendajas Medina, Pablo. 2023. "Just as Planned: Bayesianism, Externalism, and Plan Coherence." *Philosophical Imprint* 23:28. https://doi.org/10.3998/phimp.1300

Appendix A

In this appendix, we prove Lemma 1.

Proof of Lemma 1. We start by showing that (i) is continuous. Observe that (i) is a sum of terms of the form

$$\lambda(x)(w) \cdot \mathbf{I}\left[g_{\text{meta}}\left(\lambda(1), \mathsf{G}^{\mathsf{E}}(w)\right), w\right]. \tag{37}$$

Notice that $\lambda(x)(w) = P_{\mathrm{R}}(w) \cdot x + P_{\mathrm{W}}(w)(1-x)$ is a polynomial and so is continuous everywhere. Moreover, $\mathbf{I}\left[g_{\text{meta}}\left(\lambda(1), \mathsf{G}^{\mathsf{E}}(w)\right), w\right]$ is a constant. Therefore, (i) is a linear combination of continuous functions and therefore is itself continuous.

https://doi.org/10.1017/psa.2024.36 Published online by Cambridge University Press

Next we will show that (ii) is continuous at 1. To begin, observe that (ii) is a sum of terms of the form

$$\lambda(x)(w) \cdot \mathbf{I}\big[g_{\text{cond}}\big(\lambda(x), \mathsf{G}^{\mathsf{E}}(w)\big), w\big]. \tag{38}$$

Thus, to show that (ii) is continuous at 1, it suffices to show that (38) is a continuous function at 1 for all $w \in \Omega$. We have seen that $\lambda(x)(w)$ is a polynomial and so is continuous everywhere. Thus, to show that (38) is continuous at 1 it suffices to show that

$$\mathbf{I}\big[g_{\text{cond}}\big(\lambda(x), \mathsf{G}^{\mathsf{E}}(w)\big), w\big] \tag{39}$$

is continuous at 1. By our assumption that $\mathbf{I}$ satisfies Additivity, we have that $\mathbf{I}\big[g_{\text{cond}}\big(\lambda(x), \mathsf{G}^{\mathsf{E}}(w)\big), w\big]$ is equal to

$$\sum_{H \in \mathcal{P}\,\Omega} i_w^H\big[g_{\text{cond}}\big(\lambda(x), \mathsf{G}^{\mathsf{E}}(w)\big)\big]. \tag{40}$$

Fix an arbitrary $H \in \mathcal{P}(\Omega)$. To show that (39) is continuous at 1 it suffices to show that

$$f(x) = i_w^H\big[g_{\text{cond}}\big(\lambda(x), \mathsf{G}^{\mathsf{E}}(w)\big)\big] \tag{41}$$

is continuous at 1. Define $h(x)$ as follows:

$$h(x) = g_{\text{cond}}\big(\lambda(x), \mathsf{G}^{\mathsf{E}}(w)\big)(H) = \lambda(x)(H|\mathsf{G}^{\mathsf{E}}(w)). \tag{42}$$

Then $f(x) = i_w^H \circ h(x)$. By our assumption of Continuity for the local inaccuracy measure i_w^H, we know that i_w^H is a continuous function of $h(x)$. Thus, to show that $f(x)$ is continuous at 1, it suffices to show that h is continuous at 1. By the definition of $\lambda(x)(H|\mathsf{G}^{\mathsf{E}}(w))$, we have

$$h(x) = \lambda(x)(H|\mathsf{G}^{\mathsf{E}}(w)) = \frac{\lambda(x)(H \wedge \mathsf{G}^{\mathsf{E}}(w)}{\lambda(x)\big(\mathsf{G}^{\mathsf{E}}(w)\big)} = \frac{P_{\text{R}}\big(H \wedge \mathsf{G}^{\mathsf{E}}(w)\big)x + P_{\text{W}}\big(H \wedge \mathsf{G}^{\mathsf{E}}(w)\big)(1-x)}{P_{\text{R}}\big(\mathsf{G}^{\mathsf{E}}(w)\big)x + P_{\text{W}}\big(\mathsf{G}^{\mathsf{E}}(w)\big)(1-x)}. \tag{43}$$

It follows from our assumption that $\lambda(1)(\mathsf{E}(w)) > 0$ for all $w \in \Omega$ that $\lambda(1)\big(\mathsf{G}^{\mathsf{E}}(w)\big) > 0$ for all $w \in \Omega$. Since the numerator and denominator are both continuous at 1 and the denominator is greater than zero when $x = 1$, it follows that $h(x)$ is continuous at 1. This completes the proof of Lemma 1.

Cite this article: Schultheis, Ginger. 2025. "Accurate Updating." *Philosophy of Science* 92 (2):324–343. https://doi.org/10.1017/psa.2024.36