ACCURACY AND UR-PRIOR CONDITIONALIZATION

NILANJAN DAS

New York University Shanghai

Abstract.   Recently, several epistemologists have defended an attractive principle of epistemic rationality, which we shall call *Ur-Prior Conditionalization*. In this essay, I ask whether we can justify this principle by appealing to the epistemic goal of accuracy. I argue that any such accuracy-based argument will be in tension with *Evidence Externalism*, i.e., the view that agent's evidence may entail nontrivial propositions about the external world. This is because any such argument will crucially require the assumption that, independently of all empirical evidence, it is rational for an agent to be certain that her evidence will always include truths, and that she will always have perfect introspective access to her own evidence. This assumption is incompatible with *Evidence Externalism*. I go on to suggest that even if we don't accept *Evidence Externalism*, the prospects for any accuracy-based justification for *Ur-Prior Conditionalization* are bleak.

Recently, several epistemologists have defended a principle of epistemic rationality, which we shall call *Ur-Prior Conditionalization*.[1] This principle requires us to appeal to the notion of an *ur-prior*, an initial credence function that an agent has independently of all empirical evidence. The principle says the following:

> *Ur-Prior Conditionalization*. Suppose E_t is an agent's total evidence at a time t, and p_t is the posterior credence function that she adopts at t. Then, epistemic rationality requires that there be some rationally permissible ur-prior μ such that, for any proposition H,
>
> $$p_t(H) = \mu(H|E_t) = \frac{\mu(H \cap E_t)}{\mu(E_t)} \text{ (provided } \mu(E_t) > 0).$$

Ur-Prior Conditionalization is attractive for two reasons. On the one hand, it yields a theory of belief-revision which avoids certain difficulties that arise for *Bayesian Conditionalization*. On the other hand, it is a natural formal analogue of the widely accepted *Principle of Total Evidence*, the principle which says that an agent should match her credences to the degrees of evidential support provided by her total evidence.

In this essay, I explore the connection between *Ur-Prior Conditionalization* and the *accuracy-first approach to epistemology*. According to the accuracy-first approach to epistemology, the sole source of value for our credences is *gradational accuracy* (i.e., proximity to the truth), and constraints of epistemic rationality can be justified solely by appeal to accuracy-based considerations. Some writers—e.g., Oddie (1997), Greaves & Wallace (2006), Easwaran (2013), and Briggs & Pettigrew (2016)—have tried to justify *Bayesian*

[1] Constraints of this kind have been discussed by Williamson (2000), Meacham (2008), Titelbaum (2014), and Hedden (2015).

Conditionalization by appealing to the epistemic goal of accuracy. The question I want to ask is this: Can a similar accuracy-based argument be given for *Ur-Prior Conditionalization*?

I argue that any such accuracy-based argument will be in tension with *Evidence Externalism*.

> *Evidence Externalism*. An agent's evidence may entail nontrivial propositions about the external world.[2]

Evidence Externalism is incompatible with two different conceptions of evidence. First, it is incompatible with a *nonpropositionalist* conception of evidence on which an agent's evidence isn't a proposition or a set of propositions, and therefore doesn't entail any propositions at all. An example of such a view would be an account on which an agent's evidence consists of mental objects such as sense-data, or external objects such as fingerprints. Second, *Evidence Externalism* is incompatible with a *propositionalist*, but *internalist* conception of evidence on which an agent's evidence is a proposition or a set of propositions, but it only entails propositions about the agent's nonfactive mental states, e.g., her phenomenal states.[3]

Any accuracy-based argument for *Ur-Prior Conditionalization* will be in tension with *Evidence Externalism*. Why? Any such argument will crucially require an assumption: namely, that independently of all empirical evidence, epistemic rationality requires every agent to be certain that her total evidence will always entail truths, and that she will always have perfect introspective access to her total evidence. This assumption is incompatible with an externalist conception of evidence. Moreover, even if we don't accept *Evidence Externalism*, the prospects for any accuracy-based justification for *Ur-Prior Conditionalization* are bleak.

This has two consequences. On the one hand, it shows that defenders of *Ur-Prior Conditionalization* cannot rely solely on considerations of accuracy to justify the principle that they wish to defend. On the other hand, it shows that there is a tension between the accuracy-first approach to epistemology and the *Principle of Total Evidence*. Given the appeal of this principle, that's bad news for accuracy-first epistemology.

Here is the plan for this essay. I begin by motivating *Ur-Prior Conditionalization* (§1–§2). Then, I lay out the normative and evaluative assumptions that defenders of accuracy-first epistemology make (§3). I then sketch a formal framework within which we can investigate the question of whether there could be an accuracy-based argument for *Ur-Prior Conditionalization* (§4). Next, I show that under certain assumptions about evidence, rationality and measures of accuracy, updating one's credence by conditionalizing one's ur-prior on one's total evidence maximizes expected accuracy by lights of that ur-prior (§5). But this argument for *Ur-Prior Conditionalization* does not succeed if we accept *Evidence Externalism* (§6). Moreover, even if we reject *Evidence Externalism* or the other assumptions about rationality or measures of accuracy, it will still be difficult to defend *Ur-Prior Conditionalization* by appealing solely to accuracy (§7). Finally, I explore some of the consequences that this argument has for defenders of *Ur-Prior Conditionalization* and accuracy-first epistemology (§8).

---

[2] McDowell (1995, 2011), Williamson (2000), and Goldman (2009) are three prominent defenders of *Evidence Externalism*.

[3] For a recent nonpropositionalist account of evidence, see Conee & Feldman (2008). For a defense of an internalist account of evidence, see Silins (2005).

§1. Motivation I: advantages over Bayesian Conditionalization. According to Bayesian orthodoxy, the following constraint is true.

> *Bayesian Conditionalization.* Suppose E is the strongest proposition that an agent learns in a particular situation. If p_{new} is the posterior credence function that she adopts after learning E, and p_{old} is her credence function before she learns E, then epistemic rationality requires that, for any proposition H,
>
> $$p_{new}(H) = p_{old}(H|E) = \frac{p_{old}(H \cap E)}{p_{old}(E)} \text{ (provided } p_{old}(E) > 0).$$

Though many arguments have been offered for *Bayesian Conditionalization*,[4] it does not yield an adequate theory of rational belief-revision. In particular, it can't handle scenarios of *evidence loss* very well.[5]

To see why, let us note two features of *Bayesian Conditionalization*.

> *Independence.* Bayesian Conditionalization entails that, no matter what other evidence F the agent learns after learning E, her rational credence in E should remain 1.[6]
>
> *Zero Measure Events.* Bayesian Conditionalization is silent on how an agent should revise her credence in a proposition H when she receives some evidence E such that her prior credence function p_{old} assigns probability 0 to E.[7]

Since *Independence* is true, *Bayesian Conditionalization* does not adequately handle cases of *forgetting*. And, because of *Zero Measure Events*, *Bayesian Conditionalization* fails to make predictions about cases where an agent gains new *self-locating information*, i.e., information about herself or her spatiotemporal location, which is incompatible with her previous evidence. Both kinds of cases are cases of *evidence loss*. In the case of forgetting, the agent loses her previous evidence. In the case of self-locating information change, the agent receives new information that is incompatible with her previous evidence,

---

[4] For example, Teller (1973) offers a Dutchbook argument for *Bayesian Conditionalization*. Williams (1980) uses the Principle of Minimum Information to defend it. Van Fraassen (1989, 1999), appeals to his Reflection Principle and to certain symmetry considerations to argue for it. More recently, Oddie (1997), Greaves & Wallace (2006), Easwaran (2013), and Briggs & Pettigrew (2016) have offered accuracy-based arguments for *Bayesian Conditionalization*.

[5] For other, more controversial, problems for *Bayesian Conditionalization*, see Meacham (2016).

[6] This is easy to see:

$$p_{new}(E|F) = \frac{p_{new}(E \cap F)}{p_{new}(F)}$$
$$\Leftrightarrow p_{new}(E|F) = \frac{p_{old}(E \cap F|E)}{p_{old}(F|E)}$$
$$\Leftrightarrow p_{new}(E|F) = \frac{p_{old}(F|E)}{p_{old}(F|E)}$$
$$\Leftrightarrow p_{new}(E|F) = 1.$$

[7] For discussion of this problem more generally, see Hajek (2003).

and therefore must get rid of a part of her previous evidence in order to accommodate that new information.

1.1. Forgetting.   Let us begin with examples of forgetting.[8] Sometimes, we forget information over time. For example, today my evidence entails that it was raining in New York on January 1, 2017, so I am certain that it was raining in New York on January 1, 2017. But one year later, I may forget that information; so, my evidence won't entail that claim. As a result, it seems rationally permissible for me to be less confident that it was raining in New York on January 1, 2017.

Now, *Bayesian Conditionalization* yields the wrong result in these cases. Since *Independence* is true, it entails that once I have updated rationally on the proposition that it was raining in New York on January 1, 2017, my unconditional credence in that proposition should remain the same no matter what evidence I get. This means that I can't ever be unsure of that proposition. Thus, the theory of rational belief-revision that *Bayesian Conditionalization* yields is inadequate.

1.2. Self-locating information.   Let us now consider cases where an agent gains new *self-locating information*, i.e., information about herself or her spatiotemporal location, which is incompatible with her previous evidence.[9] Following a suggestion by Quine (1969), Lewis (1979) proposed a framework within which self-locating information can be represented as *centered propositions*, i.e., as sets of *centered worlds* $\langle w, c \rangle$ where w is a possible world and the center c is an ordered pair $\langle i, t \rangle$ containing a subject i and a time t. Here are some examples. The centered proposition that it is *now* 12.00 p.m. on a Wednesday is the set of all and only centered worlds $\langle w, \langle i, t \rangle \rangle$ where the time t at the center is 12.00 p.m. on a Wednesday. The centered proposition that I am NN is the set of all and only centered worlds $\langle w, \langle i, t \rangle \rangle$ where the subject i at the center is NN. Moreover, within this framework, even non-self-locating information can be represented as centered propositions. For example, the information that NN sleeps at 12.00 p.m. on November 30, 2016 is a centered proposition which contains all and only centered worlds $\langle w, \langle i, t \rangle \rangle$ where the world w is a world in which NN sleeps at 12.00 p.m. on November 30, 2016. We are going to assume that an agent's total evidence at any time in any world is a centered proposition.

Take a concrete example. Suppose I am looking at a clock, and the time it shows is 12.00 p.m. So, I update on this self-locating information, and become certain that it is *now* 12.00 p.m. I assign credence 1 to the set of centered worlds $\langle w, \langle i, t \rangle \rangle$ where t is 12.00 p.m. But then a minute later, the time the clock shows is 12.01 p.m. Once again, I receive evidence that it is *now* 12.01 p.m. This evidence is incompatible with the claim that I was earlier certain about.

Since I earlier assigned credence 1 to the set of centered worlds $\langle w, \langle i, t \rangle \rangle$ where t is 12.00 p.m., I assigned credence 0 to the set of centered worlds $\langle w, \langle i, t \rangle \rangle$ where t is 12.01 p.m. As *Zero Measure Events* shows, *Bayesian Conditionalization* is silent on how I should update my beliefs in this case. So, if *Bayesian Conditionalization* captures the only rational constraint on belief-revision, I am permitted by rationality to update my credences any way I like in this scenario. This, obviously, is bad. Once again, this reveals

---

[8] For discussions of this problem, see Skyrms (1983), Talbott (1991), Williamson (2000), and Hedden (2015).

[9] For discussion, see Elga (2000), Halpern (2004), Meacham (2008), Bradley (2011), Moss (2012), Schwarz (2012), and Titelbaum (2014).

that *Bayesian Conditionalization* does not give us a sufficiently comprehensive theory of rational belief-revision.[10]

1.3. Ur-Prior Conditionalization.   According to *Ur-Prior Conditionalization*, if p_t is the posterior credence function that an agent adopts at t and E_t is the agent's total evidence at t, then epistemic rationality requires that, for any centered proposition H, $p_t(H) = \mu(H|E_t)$ (provided $\mu(E_t) > 0$), where μ is a rationally permissible ur-prior. Before I say how *Ur-Prior Conditionalization* handles the problems that arose for *Bayesian Conditionalization*, let me make a few remarks about the *ur-prior*.

Now, an *ur-prior* of an agent is an *initial credence function* that an agent has independently of receiving any empirical information about the world or her location in the world.[11] A *rationally permissible* ur-prior is an initial credence function that is rationally permissible for an agent to have independently of receiving any empirical information about the world or her location in the world. I shall assume that there is at least one probability measure μ which counts as an ur-prior that is rationally permissible for any agent to adopt.

Let us now address the case of forgetting. At t_1, my total evidence is a centered proposition E_1 which contains only those centered worlds $\langle w, \langle i, t \rangle \rangle$ where it rains in New York on January 1, 2017. But then, one year later, at t_2, my total evidence is E_2 which includes some centered worlds $\langle w, \langle i, t \rangle \rangle$ where it doesn't rain in New York on January 1, 2017. This reflects the fact that I have lost the information that it was raining in New York on January 1, 2017. Letting *Rain* be the centered proposition that it was raining in New York on January 1, 2017, it may indeed be the case that $\mu(\sim Rain|E_2) > 0$. In that case,

---

[10] Even though the problem here arises due to *Zero Measure Events*, it is somewhat different from other problems that arise due to this feature of *Bayesian Conditionalization*. Arguably, there are cases where an agent learns some information E to which she previously rationally assigned credence 0, but E isn't incompatible with the evidence that the agent had previously. For example, suppose I learn that a dart with a point-sized tip will be thrown at a straight line representing the [0,1] interval. I have no more reason to think that it will land on any one point on the line than on any other. Since there are uncountably many points on the line, I rationally assign credence 0 to the proposition that it will land on the 1/2 mark on the line. Then, I am told that the dart has landed on that point. How should I now change my credences? Due to *Zero Measure Events*, *Bayesian Conditionalization* is once again silent about this case.

However, in response to this problem, a defender of *Bayesian Conditionalization* may be able to offer a solution that won't be available in cases where an agent gains self-locating evidence that is incompatible with her previous evidence. For example, some think that in the scenario described above, one ought to assign nonstandard probabilities (i.e., hyperreal numbers) to the proposition that the dart will land on the 1/2 mark. For sympathetic philosophical discussion, see McGee (1994), and for dissent, see Easwaran (2014). If this proposal works, then one can use the standard ratio formula of conditional probability to generate predictions about the agent's posterior credences even in such a scenario. However, this strategy isn't available in the scenario where one gains evidence that was incompatible with one's previous evidence: since one updated by conditionalizing on one's previous evidence, one had to assign credence 0 to the evidence that one came to learn later. What this shows is that things are much worse in this latter scenario than in other cases. *Bayesian Conditionalization* requires us to intersect the set of possible worlds compatible with our previous evidence with the new evidence we gain, and then to redistribute our credences over that new set of possibilities in a certain way. However, in a scenario where an agent gains contradictory evidence, intersecting the previous evidence with the new evidence results in the empty set, and no coherent credence function can be defined over that set. Many thanks to an anonymous referee for comments here.

[11] For various interpretations of the notion of ur-prior, see Meacham (2016).

$\mu(Rain|E_2) < \mu(Rain|E_1) = 1$. So, *Ur-Prior Conditionalization* explains why I should lower my confidence that it rained in New York on January 1, 2017.

Ur-Prior Conditionalization also gives us a nice way of modeling scenarios where an agent receives new self-locating evidence that is incompatible with her earlier evidence. When at 12.00 p.m., I learn that it is now 12.00 p.m., my total evidence is a centered proposition E_1 such that, for any centered world $\langle w, \langle i, t \rangle \rangle \in E_1$, t is 12.00 p.m. on some day. But when at 12.01 p.m. I learn that it is now 12.01 p.m., I lose my previous evidence that the time is now 12.00 p.m. At 12.01 p.m., my total evidence is a centered proposition E_2 such that for any centered world $\langle w, \langle i, t \rangle \rangle \in E_2$, t is 12.01 p.m. on some day. Presumably, independently of all empirical evidence, I can't rule out the centered worlds in E_2. So, if I am rational, my *ur-prior* μ plausibly shouldn't assign zero probability to E_2. Therefore, since $\mu(E_2) > 0$, $\mu(.|E_2)$ will be defined. Thus, if we impose certain plausible constraints on the ur-prior, *Ur-Prior Conditionalization* will yield concrete predictions about cases where an agent receives evidence that contradicts her earlier beliefs.

Thus, *Ur-Prior Conditionalization* takes care of some of the problems that arise for *Bayesian Conditionalization*.

§2. Motivation II: the Principle of Total Evidence. Besides these advantages over *Bayesian Conditionalization*, *Ur-Prior Conditionalization* also seems to fit a widely accepted conception of epistemic rationality. According to this picture, from an epistemic standpoint, it is rationally permissible for an agent to hold a doxastic attitude if and only if that doxastic attitude is well-proportioned to her evidence. Call this view *evidentialism*.[12]

Evidentialism just says that the limits of epistemic rationality are fixed by the agent's evidence; it doesn't say *which* parts of the agent's evidence make *which* beliefs rational. Here is a natural way of precisifying the view.

> *Principle of Total Evidence* (*First Pass*). From an epistemic standpoint, an agent is rationally permitted to hold a certain doxastic attitude towards a claim P if and only if the doxastic attitude adequately reflects the degree of support P enjoys relative to the agent's *total* body of evidence.[13]

One worry about this version of the *Principle of Total Evidence* might be this. At least, according to a popular version of *epistemic permissivism*, it can be rationally permissible for an agent to hold different doxastic attitudes towards the same proposition on the basis

---

[12] Williamson (2000), p. 164, and Kelly (2008a), §2, call this thesis a "platitude," while Conee & Richard (1985) and Adler (2002) have defended it explicitly. It is worth distinguishing my use of the word 'evidentialism' from three other uses. First, some writers like Conee & Feldman (2004) take evidentialism to entail that our evidence cannot consist in anything but our mental states. Other writers, like Fantl & McGrath (2009), construe evidentialism as the view that what is rationally permissible for an agent to believe depends solely on her evidence, and not on any pragmatic factors. Finally, some writers like Rinard (2015) also use 'evidentialism' to pick out the view that there are no practical reasons for belief. My definition of 'evidentialism' is compatible with the last two of these views, but entails none of them: it is neutral on whether epistemic rationality or justification has a pragmatic component, and on whether there might be practical reasons for holding certain beliefs. However, contrary to Conee and Feldman, I will assume that our evidence consists solely of *propositions*, and cannot include mental states themselves.

[13] The *Principle of Total Evidence* has been defended most prominently by defended by Carnap (1962) and Hempel (1965). In epistemology and philosophy of science, it also has been assumed and defended by Salmon (1967), Sober (1975, 2009), Adler (1989), Williamson (2000), Davidson (2004), and Kelly (2008a, 2008b).

of the same body of evidence, because there are different standards of weighing one's evidence, i.e., what White (2005) and Schoenfield (2013) call 'epistemic standards.'[14] So, contrary to what this version of *Principle of Total Evidence* presupposes, there might not be any unique degree of evidential support that a proposition enjoys relative to a particular body of evidence. To avoid this worry, we may simply restate *Principle of Total Evidence* as follows:

> *Principle of Total Evidence* (*Second Pass*). From an epistemic standpoint, an agent is rationally permitted to hold a certain doxastic attitude towards a claim P if and only if the doxastic attitude adequately reflects the degree of support P enjoys relative to the agent's *total* body of evidence and a rationally permissible epistemic standard.

Now, under one interpretation, the degrees of evidential support that various propositions receive from any body of evidence should be represented as a probability function. Here is a way of fleshing this thought out using ur-priors. We may think of an ur-prior μ as an epistemic standard by which an agent weighs her evidence: for any centered proposition H, $\mu(H|E_t)$ is the degree of evidential support that E_t provides to H relative to the epistemic standard laid down by μ. When μ is a probability function, the degrees of evidential support that various propositions have according to μ will indeed form a probability function.

If this way of thinking of evidential support is right, then, according to the *Principle of Total Evidence*, it is rationally permissible for an agent to assign a credence of r to H relative to her total evidence E_t if and only if $\mu(H|E_t) = r$ where μ is a rationally permissible ur-prior. This is precisely what *Ur-Prior Conditionalization* says. Thus, *Ur-Prior Conditionalization* is a natural probabilistic analogue of the *Principle of Total Evidence*.

Moreover, giving up permissivism doesn't take away the appeal of *Ur-Prior Conditionalization*. For example, if we deny that there is more than one rationally permissible epistemic standard by which an agent may weigh her evidence, even then there will be *at least* one rationally permissible ur-prior which will reflect the degrees of support that various propositions enjoy relative to different bodies of evidence. So, even then, *Ur-Prior Conditionalization* will require any agent to match her credences to the degree of evidential support according to that uniquely rational ur-prior. Hence, *Ur-Prior Conditionalization* should seem attractive even to the impermissivist. In fact, some defenders of *Ur-Prior Conditionalization*—such as Williamson (2000) and Hedden (2015)—are impermissivists of this kind.

§3. **Accuracy first.**   How can we show that *Ur-Prior Conditionalization* is correct, i.e., that epistemic rationality requires us to update by conditionalizing a rationally permissible ur-prior on our total evidence?

One might think that the resources of epistemic utility theory could be useful here. Epistemic utility theory has two aims. On the one hand, it seeks to articulate a conception of epistemic value that explains what makes one doxastic state more valuable from an epistemic standpoint than another. On the other hand, it seeks to explain why certain epistemic norms have the force that they have, by showing that conforming to them is the best means towards promoting the relevant kind of epistemic value. For instance, many defenders of epistemic utility theory accept the following conception of epistemic value.

---

[14] For recent discussion of permissivism, see Horowitz (2014), Greco & Hedden (2016), Titelbaum & Kopec (forthcoming), Schultheis (forthcoming), and Schoenfield (forthcoming).

> *Credal Veritism*. The only source of value for *credences* or *degrees of belief* that is relevant to their epistemic status is their gradational accuracy, where the gradational accuracy of a credence in a true proposition is higher when the credence is closer to 1, while the gradational accuracy of a false proposition is higher when the credence is closer to 0.[15]

Then, they show that various epistemic norms, such as Probabilism, i.e., the norm of having probabilistically coherent credences, the Principal Principle, i.e., the norm of conforming one's credences to the objective chances under certain circumstances, etc., can be shown to cohere with the instrumentally rational pursuit of accuracy. This version of epistemic utility theory is what I shall call *accuracy-first epistemology*.

Defenders of accuracy-first epistemology have offered arguments for *Bayesian Conditionalization*. For example, Greaves & Wallace (2006) have argued for *Bayesian Conditionalization* by appealing to expected accuracy: they have shown that, under certain circumstances, the expected accuracy of revising one's credences by conditionalizing is greater than revising one's beliefs according to any other rule. So, if we think that instrumental rationality requires an agent to maximize expected value, then, from an epistemic standpoint, it is instrumentally rational only to update by conditionalizing under those circumstances. Similarly, Briggs and Pettigrew (ms.) have argued for *Bayesian Conditionalization* by appealing to accuracy-dominance: under certain circumstances, failing to conditionalize leaves an agent vulnerable to a sure loss of accuracy. So, if we think that instrumental rationality requires an agent to avoid sure losses, then, from an epistemic standpoint, it is instrumentally rational only to update by conditionalizing under the relevant circumstances.

A natural question, therefore, is this: Can we offer an accuracy-based argument for *Ur-Prior Conditionalization*? I claim that this will be difficult. In §5, I say that, if we make certain assumptions about evidence and rationality, we can indeed come up with an accuracy-based argument for *Ur-Prior Conditionalization*. However, in §6–§7, I argue that there is good reason to reject some of these assumptions; once we reject them, the accuracy-based argument for *Ur-Prior Conditionalization* fails.

§4. A formal framework.   In order to state the accuracy-based argument for *Ur-Prior Conditionalization*, I will need some formal machinery. The formal framework that I shall introduce will involve two components. The first component consists of Kripke- or Hintikka-style relational structures, which I call *self-locating frames*, for representing the information state of any agent. The second component consists of *epistemic scoring rules*, which defenders of accuracy-first epistemology use to measure the accuracy of credence functions.

4.1. Self-locating frames.   Let a *self-locating frame* be a structure $\mathcal{F} = \langle W, I, T, \geq , S, E, \mu \rangle$.[16] First, W is a finite set of possible worlds. Second, I is a finite set of subjects or believers. T is a finite set of times. The relation $\geq$ is a binary transitive connected antisymmetric relation on T, a relation that determines a linear order over the set of times. The assumption that W, I, and T are finite might seem a little artificial: after all, it is natural to think that the sets of all epistemically possible worlds, agents, and times ought

---

[15]   See, for example, Joyce (1998, 2009), and Pettigrew (2016).

[16]   Discussions of similar frames occur in Halpern (2004) and Stalnaker (2008).

to be uncountably infinite. However, I take this to be a harmless idealization; with some additional constraints, all the results proved in this article can be generalized to frames that involve infinite sets of possible worlds, agents, and times.[17]

Next, we define the notion of a *centered world* loosely introduced earlier. Let $C = I \times T$ be the set of centers, which are ordered pairs $\langle i, t \rangle$ where $i \in I$ is a subject and $t \in T$ is a time. $W \times C$ is the set of all centered worlds $\langle w, c \rangle$ where $w \in W$ is a world and $c \in C$ is a centre. Now, the element S in the frame $\mathcal{F}$ is a subset of $W \times C$, such that, for any $\langle w, \langle i, t \rangle \rangle \in S$, the agent i exists at time t in w. A *centered proposition* is a subset of S. The power set of S, $\mathcal{P}(S)$, is the set of all centered propositions.

The evidence function $E : S \to \mathcal{P}(S)$ (equivalent to an *accessibility relation* in Kripke- or Hintikka-style relational structures) is a function that maps each centered world $\langle w, \langle i, t \rangle \rangle \in S$ to a set of centered worlds in S, which represents the agent i's *total evidence* at t in w. I will say that an agent i's total evidence at time t in a world w entails a certain centered proposition X if and only if her evidence at that world, i.e., $E(\langle w, \langle i, t \rangle \rangle)$, is a subset of X. Finally, $\mu : \mathcal{P}(S) \to [0, 1]$ is a rationally permissible *ur-prior*, a probability measure that takes any centered proposition $X \subseteq S$ to a real number $\mu(X)$ between 0 and 1 (inclusive). For simplicity, for any centered world $s \in S$, I will write $\mu(\{s\})$ as $\mu(s)$.

To illustrate the notion of a self-locating frame, consider once again the scenario where I am looking at the clock as time passes. So, we can represent this scenario with a self-locating frame $\langle W, I, T, \geq, S, E, \mu \rangle$ where W contains just one world w. S contains just one subject i, i.e., me, T contains two times, t_{1200} and t_{1201} where $t_{1201} \geq t_{1200}$, but not vice-versa. Therefore, S contains two centered worlds, $s_{1200} = \langle w, \langle i, t_{1200} \rangle \rangle$ and $s_{1201} = \langle w, \langle i, t_{1201} \rangle \rangle$.

For simplicity, we may assume that I am omniscient about all the truths about the world, and about my location in it. Since my evidence in s_{1200} rules out s_{1201}, $E(s_{1200}) = \{s_{1200}\}$. Similarly, since my evidence in s_{1201} rules out s_{1200}, $E(s_{1201}) = \{s_{1201}\}$. Figure 1 is a graph-theoretic representation of my evidence in this scenario (where there is a path from a node A to a node B if and only if the world represented by B is compatible with the agent's evidence in the world represented by A).
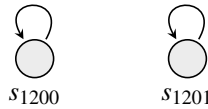

Fig. 1. My evidence in the clock example.

Finally, we may assume that, independently of getting any evidence, it is rationally permissible for me to assign nonzero initial credence to both s_{1200} and s_{1201}. So, if I comply with *Ur-Prior Conditionalization*, at 12.00 p.m., my posterior credence function is $\mu(.|E(s_{1200})) = \mu(.|\{s_{1200}\})$ and, at 12.01 p.m., my posterior credence function is $\mu(.|E(s_{1201})) = \mu(.|\{s_{1201}\})$.

4.2. Epistemic scoring rules. Typically, defenders of *Credal Veritism* measure the accuracy of a credence function using *scoring rules*.

Suppose S is a set of states (in our context, centered worlds). A credence function b defined over S maps each element s of S to a real number between 0 and 1 (inclusive),

---

[17] One can undertake this generalization much in the same way as Easwaran (2013) generalizes Greaves & Wallace's (2006) accuracy-based argument for Bayesian conditionalization.

which reflects the credence that the relevant agent in the proposition that s in fact obtains. Let B_S be the set of all such credence functions. The accuracy of a credence function b in B is measured by a scoring rule $A : B_S \times S \to [0, 1]$, which maps a credence b and a state s to the accuracy score $A(b, s)$ of b in that state s. The accuracy score of a credence function s depends on the gradational accuracy of the credences that it assigns to the elements of S. So, a credence function that assigns 1 to a state s and 0 to every other state in S will receive the maximal accuracy score of 1 in s. Similarly, a credence function that assigns 0 to a state s and 1 to every other state in S will receive the minimal accuracy score of 0 in s.

In a scenario where an agent is uncertain about which state she is in, she won't be able to figure out what the accuracy score of a credence function is. However, assuming that she is probabilistically coherent, the best she can do is use her own probabilistically coherent credence function to form an expectation of the accuracy score that the credence function has.

> *Expected Accuracy of Credence Functions.* If B_s is the set of all credence functions defined over a set of states S and $A : B_S \times S \to [0, 1]$ is a scoring rule, the *expected accuracy* of a credence function $b \in B_S$ relative to a probability function $p \in B_S$ is defined as
>
> $$Exp_p(b) = \sum_{s \in S} p(s)A(b, s).$$

In other words, the expected accuracy of a credence function b relative to a probability function p is the weighted average of the accuracy score b gets in each state s, where the weights are the probabilities p assigns to the states in S.

An important property of scoring rules is defined in terms of expected accuracy.

> *Strict Propriety.* If B_s is the set of all credence functions defined over a set of states S and $A : B_S \times S \to [0, 1]$ is a scoring rule, A is said to be *strictly proper* if and only if, for any probability function $p \in B_S$ and any credence function $b \in B_S$ distinct from p, the expected accuracy of p according to p is greater than the expected accuracy of b according to p, i.e., $Exp_p(p) > Exp_p(b)$.

Strict Propriety is supposed to capture a virtue of scoring rules, sometimes called *strict immodesty*, namely that any probabilistically coherent credence function should take itself to be uniquely optimal from an epistemic standpoint, i.e., to have the uniquely best shot at forming an accurate picture of the world. Many writers have defended some version of immodesty, and in turn the strict propriety of scoring rules.[18] I shall return to the question of strict propriety later.

In the next section, I will put the formal machinery introduced in this section to use.

§5. The argument for Ur-Prior Conditionalization. In this section, I shall state the argument for *Ur-Prior Conditionalization*. However, for our argument, we shall need two assumptions.

5.1. Assumption 1: Partitional Evidence. The first assumption is about the structure of evidence.

---

[18] For discussion, see Lewis (1971), Joyce (2009), Moss (2012), Horowitz (2014), and Pettigrew (2016).

Let us define three properties of self-locating frames. Let $\mathcal{F} = \langle W, I, T, \geq, S, E, \mu \rangle$ be a self-locating frame.

1. $\mathcal{F}$ is *reflexive* iff for any centered world $s \in S$, s is compatible with the evidence in s; formally,

$$(\forall s \in S)(s \in E(s)).$$

2. $\mathcal{F}$ is *transitive* iff for any $s, s', s'' \in S$, if s' is compatible with the evidence in s, and s'' is compatible with the evidence in s', then s'' is compatible with the evidence in s; formally,

$$(\forall s, s', s'' \in S)((s' \in E(s) \& s'' \in E(s')) \Rightarrow s'' \in E(s)).$$

3. $\mathcal{F}$ is *euclidean* iff for any $s, s', s'' \in S$, if s' is compatible with the evidence in s, and s'' is compatible with the evidence in s, then s'' is compatible with the evidence in s'; formally,

$$(\forall s, s', s'' \in S)((s' \in E(s) \& s'' \in E(s)) \Rightarrow s'' \in E(s')).$$

When a frame has all three of these properties, it is *partitional*: in such a frame, E imposes an partition on S where, for any centered state s, each $E(s)$ is a cell containing all and only those worlds in which the agent's evidence is $E(s)$.[19] We can see the frame in Figure 1 is partitional in this sense: it is reflexive, transitive, and euclidean.

Reflexivity, transitivity, and euclideanness correspond to three properties of evidence: *Global Factivity*, *Global Positive Introspection*, and *Global Negative Introspection*.[20]

> *Global Factivity.* For any centered world $\langle w, \langle i, t \rangle \rangle \in S$, if the agent i's evidence entails a centered proposition X in $\langle w, \langle i, t \rangle \rangle$, then X is true in $\langle w, \langle i, t \rangle \rangle$.

---

[19] Typically, a partitional frame is characterized as a frame that is reflexive, transitive, and *symmetric*. Here, we define the property of symmetry as follows: a frame $\mathcal{F} = \langle W, I, T, \geq, S, E, \mu \rangle$ is symmetric iff for any $s, s' \in S$, if s' is compatible with the evidence in s, then s is compatible with the evidence in s'. This characterization follows from the previous one, since a frame that is reflexive and euclidean will necessarily be transitive and symmetric. Moreover, it makes it easy to see why a partitional frame is *partitional*: we know that equivalence relations, i.e., reflexive, transitive, and symmetric relations, on a set impose a partition on the set. However, since I am interested in the relationship between *Evidence Externalism* and the accuracy-based argument for *Ur-Prior Conditionalization* (see §6), this second characterization is less helpful for my purposes. Let me explain.

Later, I show that if *Evidence Externalism* is true, either *Global Factivity* or *Global Negative Introspection* must be rejected. In particular, I investigate cases where an evidence externalist keeps *Global Factivity* but rejects *Global Negative Introspection*: since reflexivity corresponds to *Global Factivity*, I focus on reflexive frames that allow for failures of *Global Negative Introspection*. But failures of *Global Negative Introspection* don't generally coincide with failures of symmetry. They only do so when the frames we are considering are both reflexive and transitive. In fact, failures of transitivity alone could result in failures of *Global Negative Introspection* in *symmetric* frames. Therefore, if we are interested more generally in reflexive (but not necessarily transitive) frames that allow for failures of *Global Negative Introspection*, it seems better to straightforwardly talk about noneuclidean frames rather than nonsymmetric ones.

[20] This of course should be obvious from the connection between similar principles and properties of frames in standard epistemic logic.

> *Global Positive Introspection.* For any centered world $\langle w, \langle i, t \rangle \rangle \in S$, if the agent i's evidence entails a centered proposition X in $\langle w, \langle i, t \rangle \rangle$, then her evidence in $\langle w, \langle i, t \rangle \rangle$ entails *de se* that her current total evidence entails X.
>
> *Global Negative Introspection.* For any centered world $\langle w, \langle i, t \rangle \rangle \in S$, if the agent i's evidence does not entail a centered proposition X in $\langle w, \langle i, t \rangle \rangle$, then her evidence in $\langle w, \langle i, t \rangle \rangle$ entails *de se* that her current total evidence does not entail X.

We can see how these properties of evidence are reflected by my evidence in the clock example, represented by Figure 1. In each centered world, my evidence entails only truths, and my evidence entails what it does or doesn't entail. So, my evidence satisfies *Global Factivity*, *Global Positive Introspection*, and *Global Negative Introspection*.

We are now in a position to state our first assumption.

> *Partitional Evidence. Global Factivity*, *Global Positive Introspection*, and *Global Negative Introspection* are true.

5.2. Assumption 2: Rational Planning.
The second assumption is an assumption about epistemic rationality.

We start by introducing the notion of a *doxastic plan*.[21] A doxastic plan specifies what doxastic attitudes one should have in response to any possible body of evidence that one could end up with. Formally, we may think of a doxastic plan as a function that maps states to credence functions, depending on the evidence that one has in those states.

> *Doxastic Plans.* Suppose S is a set of states S such that B_S is the set of all credence functions defined over S and $E : S \rightarrow \mathcal{P}(S)$ is an evidence function that maps elements of S to subsets of S. Then, a doxastic plan $R : S \rightarrow B_S$ is a function such that for any two states $s, s^* \in S$, if $E(s) = E(s^*)$, then $R(s) = R(s^*)$.

Note why not every function from states to credence functions should count as doxastic plan. For example, consider the function R that takes an input any state and outputs a credence function that assigns 1 to all truths and 0 to all falsehoods. Now, there might be two states s and s^* where the relevant agent's evidence is the same, but certain propositions that are true in s are false in s^*. So, even though R recommends different credence functions in those two states, the relevant agent does not have any epistemic means of distinguishing the two states, and therefore cannot have any epistemic basis for adopting the distinct credence functions that R recommends in the two different states. That is why R doesn't quite count as a *plan*.

Now, we can define a notion of expected accuracy for plans.

> *Expected Accuracy of Doxastic Plans.* If B_s is the set of all credence functions defined over a set of states S and $A : B_S \times S \rightarrow [0, 1]$ is a scoring rule, the *expected accuracy* of a plan $R : S \rightarrow B_S$ relative to a probability function $p \in B_S$ is defined as
>
> $$Exp_p(R) = \sum_{s \in S} p(s)A(R(s), s).$$

---

[21] Similar notions have been discussed by Gibbard (2003), Schafer (2014), and Schoenfield (2017). What we are calling 'doxastic plans' are what Greaves & Wallace call *epistemic acts*.

Let a *cognitive decision problem* be a structure $D = \langle W, I, T, \geq, S, E, \mu, B_S, A \rangle$, where $\langle W, I, T, \geq, S, E, \mu \rangle$ is a self-locating frame, B_S is the set of all credence functions defined over S and $A : B_S \times S \to [0, 1]$ is an epistemic scoring rule. We can then define a notion of rational planning relative to cognitive decision problems.

> *Rational Planning*. Relative to a cognitive decision problem $D = \langle W, I, T, \geq, S, E, \mu, B_S, A \rangle$, it is rationally permissible for an agent to conform to a doxastic plan $R : S \to B_S$ iff, for any doxastic plan $R^* : S \to B_S$, $Exp_\mu(R) \geq Exp_\mu(R^*)$.

In other words, an agent is rationally permitted to conform to a doxastic plan relative to a cognitive decision problem iff, according to the relevant rationally permissible ur-prior, that plan maximizes expected accuracy.

Roughly, the idea underlying *Rational Planning* is this. When an agent is deciding which doxastic plan to use, she is picking a policy that allows her to respond to any evidential situation that she might find herself in. Now, in order to come up with a plan that is sufficiently flexible, she should do so from a perspective which is independent of any empirical evidence that she might have. This is because, very often, an agent's (empirical) evidence will tell her what her current evidential situation is. Hence, even if an agent might find herself *later* in a different evidential situation, she might rationally assign credence 0 to the possibility that she is *currently* in that situation. If she were to pick a plan in light of such credences, she will only be taking into account the expected accuracy of the plan in her current evidential situation. But, then, her choice of plan won't give her any guidance whatsoever when it comes to fixing her beliefs in those future evidential situations to which she currently assigns zero credence. That is why it makes sense to pick a doxastic plan independently of all empirical evidence, i.e., using solely a rationally permissible ur-prior. Using a rationally permissible ur-prior, the agent should evaluate the expected accuracy of every plan. The optimal plan by her lights is the one that maximizes expected accuracy according to her ur-prior, so she should conform to that plan. And that's precisely what *Rational Planning* says.

5.3. The argument. We can now show that conforming to the recommendations of *Ur-Prior Conditionalization* is rationally mandatory if *Partitional Evidence* and *Rational Planning* are true.

Relative to any cognitive decision problem $D = \langle W, I, T, \geq, S, E, \mu, B_S, A \rangle$, we define two kinds of plans.

> *Conditionalizing Plans*. An doxastic plan $R : S \to B_S$ is a *conditionalizing plan* iff, for any centered world s and any centered proposition H, if the credence function $p = R(s)$, $\mu(H \cap E(s)) = \mu(E(s))p(H)$.

> *Meta-Conditionalizing Plans*. For any $X \subseteq S$, let $[\mathbf{E} = X] = \{s \in S : E(s) = X\}$, i.e., the centered proposition that one's current total evidence is X. An doxastic plan $R : S \to B_S$ is a *meta-conditionalizing plan* iff, for any centered world s and any centered proposition H, if the credence function $p = R(s)$, $\mu(H \cap [\mathbf{E} = E(s)]) = \mu([\mathbf{E} = E(s)])p(H)$.

Intuitively, a conditionalizing plan is a doxastic plan which, for any centered world s, recommends the credence function $\mu(.|E(s))$, provided $\mu(E(s)) > 0$. And a meta-conditionalizing plan is a doxastic plan which, for any centered world s, recommends the credence function $\mu(.|[\mathbf{E} = E(s)])$, provided $\mu([\mathbf{E} = E(s)]) > 0$. The difference is that,

for any s, a conditionalizing plan requires the agent to conditionalize on her evidence $E(s)$ when $\mu(E(s)) > 0$; by contrast, a meta-conditionalizing plan requires the agent to conditionalize on the fact that her total evidence is $E(s)$ when μ assigns nonzero probability to that fact.

The following theorem holds.

> *Theorem 1*. Suppose $D = \langle W, I, T, \geq, S, E, \mu, B_S, A \rangle$ is a cognitive decision problem where A is a strictly proper scoring rule. Then, for any doxastic plan $R : S \to B_S$, R maximizes expected accuracy according to μ and A iff R is a meta-conditionalizing plan for μ.
>
> In other words,
>
> > (i) For any two doxastic plans $R : S \to B_S$ and $R^* : S \to B_S$, if both R and R^* are meta-conditionalizing plans for μ, then $Exp_\mu(R) = Exp_\mu(R^*)$.
> > (ii) For any two doxastic plans $R : S \to B_S$ and $R^* : S \to B_S$, if R is a meta-conditionalizing plan for μ but R^* is not, then $Exp_\mu(R) > Exp_\mu(R^*)$.[22]

For now, note that *Theorem 1* does *not* by itself vindicate *Ur-Prior Conditionalization*. It only shows that, relative to a rationally permissible ur-prior μ, all and only meta-conditionalizing plans for μ maximize expected accuracy.

In order to support *Ur-Prior Conditionalization*, we prove the following *Lemma*.

> *Lemma*. Suppose $\langle W, I, T, \geq, S, E, \mu \rangle$ is a self-locating frame. Let $\Pi = \{X \subseteq S : (\exists s \in S)(E(s) = X)\}$ be the set of all possible bodies of total evidence that one could have in S. Then, $\langle W, I, T, \geq, S, E, \mu \rangle$ is partitional iff for any centered proposition $X \in \Pi$, $[\mathbf{E} = X] = X$.

In other words, *Partitional Evidence* holds iff, for any possible evidence proposition X that an agent could end up with, $[\mathbf{E} = X] = X$.

Note why this is significant. This means that if *Global Factivity*, *Global Positive Introspection* and *Global Negative Introspection* hold, then all and only *meta-conditionalizing plans* for a rationally permissible ur-prior μ are *conditionalizing* plans for μ. As a result, conditionalizing plans will maximize expected accuracy. More formally, *Theorem 1* and *Lemma* immediately entail the following corollary.

> *Corollary 1*. Suppose $D = \langle W, I, T, \geq, S, E, \mu, B_S, A \rangle$ is a cognitive decision problem where $\langle W, I, T, \geq, S, E, \mu \rangle$ is a partitional self-locating frame and A is a strictly proper scoring rule. Then, for any doxastic plan $R : S \to B_S$, R maximizes expected accuracy according to μ and A iff R is a conditionalizing plan for μ.
>
> In other words,
>
> > (i) For any two doxastic plans $R : S \to B_S$ and $R^* : S \to B_S$, if both R and R^* are conditionalizing plans for μ, then $Exp_\mu(R) = Exp_\mu(R^*)$.
> > (ii) For any two doxastic plans $R : S \to B_S$ and $R^* : S \to B_S$, if R is a conditionalizing plan for μ but R^* is not, then $Exp_\mu(R) > Exp_\mu(R^*)$.

---

[22] All proofs are given in the Appendix.

The upshot is this. *Corollary 1* shows that when one's total evidence satisfies *Global Factivity*, *Global Positive Introspection* and *Global Negative Introspection*, a conditionalizing plan for a rationally permissible ur-prior has greater expected accuracy than any nonconditionalizing plan according to a strictly proper scoring rule and the relevant ur-prior. By *Partitional Evidence* and *Rational Planning*, therefore, it follows that it is rationally permissible for an agent to conform to a certain doxastic plan iff it is a conditionalizing plan relative to a rationally permissible ur-prior. This, in turn, justifies *Ur-Prior Conditionalization*.

Theorem 1 is similar to a theorem that Schoenfield (2017) proves in her discussion of *Bayesian Conditionalization*. Schoenfield (2017) proves that the updating plan recommended by *Bayesian Conditionalization* does not in general maximize expected accuracy; what maximizes expected accuracy is a different updating plan, which, in any scenario where the strongest proposition that an agent has learnt is E, requires her to conditionalize on the fact that the strongest proposition she has learnt is E. This plan coincides with the plan recommended by *Bayesian Conditionalization* if and only if her future learning experience is what Greaves & Wallace (2006) call an *experiment*. Roughly speaking, a learning experience is an experiment just in case the agent is antecedently certain that for any proposition E that might be the strongest proposition she learns, E is true if and only if E is the strongest proposition that she learns. In this respect, Schoenfield's argument is similar to my own. I have shown that what maximizes expected accuracy relative to an agent's ur-prior isn't the plan recommended by *Ur-Prior Conditionalization*, but rather meta-conditionalizing plans which require the agent to conditionalize on the claim that her total evidence is E whenever her total evidence is E. This plan coincides with the plan recommended by *Bayesian Conditionalization* if *Partitional Evidence* is true. And, according to *Lemma*, *Partitional Evidence* entails that for any proposition E that might be an agent's total evidence, E is true if and only if E is an agent's total evidence.

Despite these similarities, there is an important difference between Schoenfield's result and mine. In her results, Schoenfield is concerned solely with scenarios where an agent either merely gains some new evidence that is compatible with her previous evidence, or doesn't learn anything at all without losing any information. In particular, Schoenfield doesn't address cases of *information loss*, e.g., cases where an agent forgets information, or cases where she gains self-locating information that is incompatible with her previous evidence and therefore must get rid of some evidence that she previously had. By contrast, the framework within which *Theorem 1* is formulated is much more general in scope: it is compatible with scenarios of this kind.

This is important. In recent years, philosophers have discussed numerous counterexamples to *Bayesian Conditionalization* that involve scenarios precisely of this kind.[23] Despite these counterexamples, many think that some restricted version of *Bayesian Conditionalization* must still be true.[24] Now, *Theorem 1* might be helpful in articulating such a properly restricted version of *Bayesian Conditionalization*. By examining cases where the

---

[23] See footnotes 8 and 9.

[24] See, for instance, Halpern (2004), Meacham (2008), and Titelbaum (2014). Neither Halpern nor Meacham favor *Ur-Prior Conditionalization*, but they favor another rule called 'compartmentalized conditionalization' which yields a version of *Bayesian Conditionalization* restricted to non-self-locating propositions. But Titelbaum favors a principle called 'generalized conditionalization', which is entailed by *Ur-Prior Conditionalization* and yields a version of *Bayesian Conditionalization* restricted to scenarios where an agent merely gains information but loses nothing. I suspect that besides these two restricted versions of *Bayesian Conditionalization*,

recommendations of *Bayesian Conditionalization* come apart from the recommendations of meta-conditionalizing plans, the accuracy-first epistemologist may be able to come up with a general characterization of the conditions under which updating one's credence in a certain proposition or a certain class of propositions according to *Bayesian Conditionalization* maximizes expected accuracy according to one's ur-prior. This in turn would be useful for assessing various suitably restricted versions of *Bayesian Conditionalization* from the accuracy-first standpoint. Thus, since *Theorem 1* covers a wider range of scenarios than Schoenfield's result, it promises to shed more light on the conditions under which *Bayesian Conditionalization* offers the right advice from an accuracy-first standpoint.[25]

§6. The externalist objection. In this section, I argue that the defender of *Evidence Externalism* cannot accept the accuracy-based argument for *Ur-Prior Conditionalization* presented in the last section. I do so by showing that *Partitional Evidence* is in tension with *Evidence Externalism*.

6.1. Externalism and Partitional Evidence. Amongst *Global Factivity*, *Global Positive Introspection*, and *Global Negative Introspection*, *Global Factivity* seems to be the least controversial.

According to *Global Factivity*, an agent's evidence only entails truths. First of all, there are plenty of arguments in favor of *Global Factivity*.[26] Moreover, it is worth pointing out that *Ur-Prior Conditionalization* would be extremely difficult to justify from an accuracy-first perspective without *Global Factivity*; for, if our evidence were to entail falsehoods, then conditionalizing on it would give rise to inaccurate credences. Hence, it may indeed be instrumentally rational for us to avoid conditionalizing on our total evidence. This connection between *Ur-Prior Conditionalization* and *Global Factivity* is so obvious that there doesn't seem to be anything theoretically interesting about calling *Ur-Prior Conditionalization* into question by rejecting *Global Factivity*. Therefore, let us grant that *Global Factivity* is true.

What about *Global Positive Introspection* and *Global Negative Introspection*? There is a certain conception of evidence on which both these introspection principles might seem quite natural. According to a phenomenalist or Cartesian picture of evidence, an agent's evidence consists only of propositions concerning her current phenomenal states, i.e., propositions about *what it's like* for her at that time. Some think that an agent cannot be misled about such states and their absence: necessarily, if such states obtain, the agent learns by introspection that they do, and if they don't obtain, the agent learns by introspection that they don't. On this picture, therefore, when the agent's evidence includes (or doesn't include) a certain proposition, her evidence entails that her evidence includes (or doesn't include) that proposition. Therefore, both *Global Positive* and *Global Negative Introspection* are true.

However, the combination of the Cartesian picture of evidence with *Ur-Prior Conditionalization* pushes us toward skepticism about the external world. For the Cartesian, we only

there are plenty of other versions of *Bayesian Conditionalization* which are weaker than the Halpern–Meacham proposal and stronger than the Titelbaum one.

[25] Of course, I don't intend this to be taken as a criticism of Schoenfield's result, since she is concerned solely with criticizing Greaves & Wallace's (2006) accuracy-based argument for *Bayesian Conditionalization* which is formulated within a similar framework.

[26] For arguments in favour of *Global Factivity*, see Williamson (2000), Littlejohn (2012), Byrne (2013). For dissent from *Global Factivity*, see Joyce (2004), Goldman (2009), and Leite (2013).

ever have conclusive evidence for centered propositions about our current phenomenal states. Suppose at a certain point of time t, my total evidence is E_t, which is the set of all centered worlds where I am in the same phenomenal states as the ones that I am in at t. If I am to rationally believe any claim H about the external world by conditionalizing my ur-prior μ on E_t, $\mu(H|E_t)$ must be considerably higher than $\mu(\sim H|E_t)$. This means that $\mu(H \cap E_t)$ must be considerably higher than $\mu(\sim H \cap E_t)$. Now, recall, E_t is a proposition solely about my phenomenal states. And let H be the proposition that I have hands. So, if I am to be rationally very confident that I have hands on my total evidence, then, independently of all empirical evidence, I have to be considerably more confident that I have hands and am in the relevant phenomenal states, than that I am a handless brain in a vat and am in the relevant phenomenal states. How could I have nonempirical evidence for taking one of these contingent hypotheses to be much more likely than the other? If I don't any nonempirical reason to favor the first hypothesis over the latter, I cannot be rationally confident that I have hands. Thus, we are led to skepticism about the external world.[27]

This might motivate us to accept the following, more natural conception of evidence.

> *Evidence Externalism.* An agent's evidence may include nontrivial propositions about the external world.[28]

However, if both *Global Factivity* and *Evidence Externalism* are true, then *Global Negative Introspection* cannot be saved.

Suppose I am looking at a white wall that is lit up with red light, but I have no reason to think that this is the case. Since I am undergoing an experience as of there being a red wall before me, I have strong misleading evidence for thinking that the wall before me is red. If *Evidence Externalism* is correct, then, plausibly, I can gain conclusive evidence about the external world from my veridical and reliable perceptual experiences. So, when I have strong misleading evidence for thinking that the wall is red, I may have strong evidence for thinking that I have conclusive perceptual evidence that the wall is red; for I have no reason to suspect that my perceptual experience is unreliable or nonveridical. Thus, my evidence won't entail that my evidence doesn't entail that the wall is red. However, by *Global Factivity*, my evidence won't entail that the wall is red, because that claim is false. Therefore, *Global Negative Introspection* will fail.[29]

---

[27] I will consider some responses to this argument in §7.

[28] Typical examples of *Evidence Externalism* include Williamson's (2000) $E = K$ thesis, McDowell's (2011) view that when one undergoes a veridical perception, one's evidence includes the proposition that one sees that such-and-such is the case, and Goldman's (2009) view that one's evidence includes the deliverances of reliable noninferential cognitive processes. Besides the threat of skepticism, there may be other reasons for accepting *Evidence Externalism*. For example, one might think that as a conception of evidence, *Evidence Externalism* is easier to reconcile with the notion of evidence that we use in legal or scientific contexts, where we treat facts about the external world as our evidence. However, several philosophers have treated the threat of skepticism to be the strongest reason for accepting *Evidence Externalism*: see McDowell (1982, 1995, 2011), Williamson (2000), Chapters 8 and 9, Neta & Pritchard (2007), and Lasonen-Aarnio (2013).

[29] For similar complaints about the negative introspection principle about knowledge, see Hintikka (1962), p. 106, Williamson (2000), pp. 23–27, and Stalnaker (2009), p. 400. Some externalists like Goldman (2009) don't accept *Global Factivity*. However, Goldman might still reject *Global Negative Introspection*. On his view, it is possible for an agent to have misleading evidence about the reliability of a cognitive mechanism; so, an agent may reasonably take her evidence to entail a certain proposition when it in fact doesn't entail it.

More generally, the idea is this. Even when a claim P about the external world is false, an agent may have strong misleading evidence for thinking that P is true. If *Evidence Externalism* is correct, the agent may in such a scenario have strong misleading evidence for thinking that P is part of her evidence (provided that she also thinks that other conditions for P to be part of her evidence are satisfied). However, by *Global Factivity*, the agent's evidence cannot entail P. Therefore, *Global Negative Introspection* fails.

It is worth noting that some evidence externalists also take *Global Positive Introspection* to be false. Consider, for example, what Williamson (2000) calls the $E = K$ thesis, i.e., the thesis that all and only known claims are part of an agent's evidence. If this view is correct, then *Global Positive Introspection* entails that if what an agent knows entails X, then what she knows entails that what she knows entails X. This is questionable for the same reasons that cast doubt on the KK principle, i.e., the principle that if an agent knows a claim, she is in a position to know that she knows it. Since knowledge requires reliability, we might think that an agent can reliably believe a claim, without being able to reliably determine that she reliably believes it; if so, she can know without being able to know that she does.[30] Other writers, however, have resisted this argument.[31] Therefore, *Evidence Externalism* need not be straightforwardly incompatible with *Global Positive Introspection*.

The upshot is that *Evidence Externalism* is incompatible with *Partititial Evidence*: if both *Global Factivity* and *Evidence Externalism* are true, then *Global Negative Introspection* will fail. In the remainder of this section, I will show that if we reject *Partitional Evidence* on externalist grounds, then we cannot justify *Ur-Prior Conditionalization* by appealing to expected accuracy.

6.2. An example. Let us start with a scenario where *Global Factivity* and *Global Positive Introspection* are true, but *Global Negative Introspection* fails.

> *Red Wall.* At t_1, I know who I am and what time it is, but don't know what color the wall in a certain room is. However, at t_2 I will enter the room and look at the wall, while also learning that the time is t_2. There are two possibilities: either the wall will be red and lit up with normal light, or the wall will be white and lit up with trick red lighting. If it is red and the lighting conditions are normal, I will learn that it is red. If it is white but lit up with trick lighting, then I won't learn that the wall is red.

Let us formally represent the scenario with a self-locating frame $\langle W, I, T, \geq, S, E, \mu \rangle$.

Here, W consists of two worlds: the world r where the wall is red, and the world w where the wall is white. Let I include just one agent i, which is me. Let T include two times t_1 and t_2, where $t_2 \geq t_1$ but not vice-versa. Finally, $S = W \times (I \times T) = \{s_{r1}, s_{r2}, s_{w1}, s_{w2}\}$, where $s_{r1} = \langle r, \langle i, t_1 \rangle \rangle$ is the centered world where the wall is red and the time is t_1, $s_{r2} = \langle r, \langle i, t_2 \rangle \rangle$ is the centered world where the wall is red and the time is t_2, $s_{w1} = \langle w, \langle i, t_1 \rangle \rangle$ is the centered world where the wall is white and the time is t_1, and, finally, $s_{w2} = \langle w, \langle i, t_2 \rangle \rangle$ is the centered world where the wall is white and the time is t_2.

Finally, we can describe my evidence in each world. At t_1, I know that the time is t_1, but I don't know which world I am in. Therefore,

---

[30] For such complaints against the KK principle, see Alston (1980), pp. 140–141, Williams (1991), p. 96, Antony (2004), p. 12, Dretske (2004), §2, and Williamson (2000).

[31] See, for example, Stalnaker (2006, 2009, 2015), Greco (2014), and Das & Salow (2018).

(1) $E(s_{r1}) = E(s_{w1}) = \{s_{r1}, s_{w1}\}$.

Since in r at t_2, I learn that the time is t_2, and the wall is red,

(2) $E(s_{r2}) = \{s_{r2}\}$.

Since in w at t_2, I learn that the time is t_2, but don't learn that the wall is red,

(3) $E(s_{w2}) = \{s_{r2}, s_{w2}\}$.
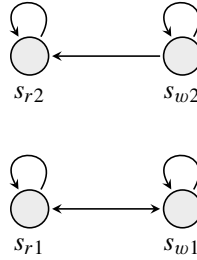
So, we can represent this scenario as follows:

Fig. 2. My evidence in *Red Wall*.

We can see why this frame is not partitional: it is noneuclidean because even though $s_{w2} \in E(s_{w2})$ and $s_{r2} \in E(s_{w2})$, $s_{w2} \notin E(s_{r2})$. Thus, *Global Negative Introspection* fails.

Finally, we take on board a minimal assumption about the ur-prior μ.

(4) For any $s \in S$, $\mu(s) > 0$.

This corresponds to the constraint of *regularity*, namely that the ur-prior assigns nonzero probability to every centered world in S.

Now, consider the cognitive decision problem $\langle W, I, T, \geq, S, E, \mu, B_S, A \rangle$.

We call a scoring rule $A : B_S \times S \to [0, 1]$ *truth-directed* iff, for credence functions $b, b^* \in B_S$, if, the credences assigned by b are uniformly closer to the truth-values of the relevant propositions than the credences assigned by b^*, then according to A, the accuracy of b^* is less than that of b. More formally, the notion of truth-directedness can be defined as follows.

> *Truth-Directedness.* Let χ_s be the function such that, for any $s^* \in S$, $\chi_s(s^*) = 1$ if $s = s^*$ and $\chi_s(s^*) = 0$ if $s \neq s^*$. Then, A is *truth-directed* iff, for any $s \in S$ and any two credence functions $b, b^* \in B_S$, if
>
> (i) $|b(s^*) - \chi_s(s^*)| \leq |b^*(s^*) - \chi_s(s^*)|$, for all $s^* \in S$ and
> (ii) $|b(s^*) - \chi_s(s^*)| < |b^*(s^*) - \chi_s(s^*)|$, for some $s^* \in S$,
>
> then $A(b^*, s) < A(b, s)$.

From an accuracy-first standpoint, truth-directedness seems like a natural constraint on scoring rules, and some version of it is accepted by many.[32] We shall assume that in the decision problem $\langle W, I, T, \geq, S, E, \mu, B_S, A \rangle$, the scoring rule A is truth-directed.

Let us now see why in this decision problem, conditionalizing plans don't maximize expected accuracy. Suppose R_{Con} is a conditionalizing plan for μ, and R_{Meta} is a meta-

---

[32] See, for instance, Joyce (2009). See also Konek (forthcoming) and Mayo-Wilson & Wheeler (2016) for restricted versions of this principle.

conditionalizing plan for μ. We can now show that the expected accuracy of R_{Con} according to any truth-directed rule in this case is less than that of R_{Meta}.

Note that, for any $s \in S$ other than s_{w2}, the evidence that the agent has in s coincides with the centered proposition that the agent has *that* evidence. In other words, for every $s \in S$ other than s_{w2}, $[\mathbf{E} = E(s)] = E(s)$. Let me explain. When the time is t_1, the agent's total evidence at both r and w is $\{s_{r1}, s_{w1}\}$. Moreover, s_{r1} and s_{w1} are also all and only those centered worlds where the agent's current total evidence is $\{s_{r1}, s_{w1}\}$. Similarly, in s_{r2}, the agent's total evidence is $\{s_{r2}\}$. Moreover, s_{r2} is also the sole centered world where the agent's current total evidence is $\{s_{r2}\}$. As a result, the recommendations of R_{Con} and R_{Meta} will coincide in any $s \in S$ other than s_{w2}. In s_{r1} and s_{w1}, both R_{Con} and R_{Meta} will recommend $\mu(.|\{s_{r1}, s_{w1}\})$. At s_{r2}, both R_{Con} and R_{Meta} will recommend $\mu(.|\{s_{r2}\})$.

However, in s_{w2}, the agent's total evidence is $\{s_{r2}, s_{w2}\}$. But s_{w2} is the only centered world where her total evidence is $\{s_{r2}, s_{w2}\}$. So, her total evidence $\{s_{r2}, s_{w2}\}$ isn't the same as the set of centered worlds where her total evidence is $\{s_{r2}, s_{w2}\}$. That is why $[\mathbf{E} = E(s_{w2})] \neq E(s_{w2})$. As a result, the recommendations of R_{Con} and R_{Meta} will come apart in s_{w2}. At s_{w2}, R_{Con} will recommend that the agent's credence function be $\mu(.|E(s_{w2})) = \mu(.|\{s_{r2}, s_{w2}\})$, while R_{Meta} will recommend that the agent's credence function be $\mu(.|[\mathbf{E} = E(s_{w2})]) = \mu(.|\{s_{w2}\})$.

Now, intuitively, we can see that in s_{w2}, the values of $\mu(.|[\mathbf{E} = E(s_{w2})])$ are uniformly closer to the truth than that of $\mu(.|E(s_{w2}))$. This is because $\mu(.|[\mathbf{E} = E(s_{w2})])$ assigns credence 1 to s_{w2} and 0 to every other world. But, by (6), $\mu(.|E(s_{w2}))$ assigns a credence between 0 and 1 (exclusive) to both s_{r2} and s_{w2}, and 0 to s_{r1} and s_{w1}. By *Truth-Directedness*, therefore, $A(R_{Meta}(s_{w2}), s_{w2}) > A(R_{Con}(s_{w2}), s_{w2})$. For any other $s \in S$, since $R_{Meta}(s) = R_{Con}(s)$, $A(R_{Meta}(s), s) = A(R_{Con}(s), s)$. Thus, the expected accuracy of R_{Meta} will be greater than that of R_{Con}.

In *Red Wall*, therefore, we have a plausible case of introspection failure where conforming to *Ur-Prior Conditionalization* doesn't maximize expected accuracy, but conforming to a meta-conditionalizing plan does.

6.3. The worry generalized. The worry can be put more generally. *Theorem 1* and *Lemma* entail the following.

> *Corollary 2.* Suppose $D = \langle W, I, T, \geq, S, E, \mu, B_S, A \rangle$ is a cognitive decision problem where A is a strictly proper scoring rule. Then, the following are inconsistent.
>
> > (i) For any two doxastic plans $R : S \to B_S$ and $R^* : S \to B_S$, if R is a conditionalizing plan relative to μ, $Exp_\mu(R) \geq Exp_\mu(R^*)$.
> > (ii) μ is a *regular* probability function, i.e., for any $s \in S$, $\mu(s) > 0$.
> > (iii) $\langle W, I, T, \geq, S, E, \mu \rangle$ is a nonpartitional self-locating frame.

The rough thought is this. Suppose *Partitional Evidence* is false; so, there are some centered worlds where either *Global Factivity*, *Global Positive Introspection*, or *Global Negative Introspection* fail. Now, if an agent has a regular ur-prior, then her ur-prior will assign nonzero probability to some of the centered worlds where the agent either has falsehoods as part of her evidence or lacks perfect access to her own evidence. According to *Lemma*, that means that conditionalizing plans won't coincide with meta-conditionalizing plans. According to *Theorem 1*, this implies that conditionalizing plans for an ur-prior μ won't maximize expected accuracy relative to μ.

This shows that a defender of *Evidence Externalism* should not accept our accuracy-based argument for *Ur-Prior Conditionalization*. If both *Global Factivity* and *Evidence Externalism* are true, then *Global Negative Introspection* will be false: there will be some centered worlds where an agent's evidence doesn't entail a proposition but her evidence doesn't entail that her evidence doesn't entail it. Hence, *Partitional Evidence* will be false. Then, according to *Corollary 2*, conditionalizing plans won't maximize expected accuracy by lights of a regular ur-prior. Now, it may indeed be rationally permissible for an agent to have a regular ur-prior defined over a finite possibility space. According *Rational Planning*, therefore, there indeed may be rational agents who are not subject to the requirement of conforming to conditionalizing plans relative to their ur-priors. This tells against *Ur-Prior Conditionalization*. Thus, we have shown that there is a tension between our accuracy-based argument for *Ur-Prior Conditionalization* and *Evidence Externalism*.

Others—such as Bronfman (2014) and Schoenfield (2017)—have raised somewhat similar objections to Greaves & Wallace's (2006) accuracy-based argument for *Bayesian Conditionalization*. However, the objection presented in this section is different in two respects. First of all, both Bronfman (2014) and Schoenfield (2017) show that in cases where an agent assigns nonzero probability to the possibility that she will learn some evidence E in the future without learning that she has learnt E, conforming to *Bayesian Conditionalization* won't maximize expected accuracy. Now, when an agent assigns nonzero probability to such possibilities, she takes seriously the possibility that *Global Positive Introspection* might fail. In this respect, both Bronfman and Schoenfield rely quite heavily on Williamson's (2000) anti-luminosity and anti-KK arguments, both of which involve certain controversial margin-for-error principles and have been criticized recently for that reason.[33] In comparison, my argument is on steadier ground. I have argued that if we want to escape skepticism about the external world, we must accept *Evidence Externalism*, and if we accept *Evidence Externalism*, the accuracy-based argument for *Ur-Prior Conditionalization* cannot succeed, since either *Global Factivity* or *Global Negative Introspection* has to be false. My approach, therefore, is independent of anything as strong as the anti-luminosity and the anti-KK arguments.

Second, even if Bronfman's and Schoenfield's arguments succeed, they don't take away the force of Greaves & Wallace's (2006) accuracy-based argument for *Bayesian Conditionalization*. Textbooks of experimental design teach us how to design learning experiences carefully, so that we leave no room for deception or overlooking evidence. If an agent takes sufficient care in designing her learning experiments, she can be antecedently sure that, for any E that she might learn, E is true if and only if she also learns that the strongest proposition she learns is E. In such cases, relative to the agent's prior credence function, conforming to the updating plan recommended by *Bayesian Conditionalization* will indeed maximize expected accuracy. By contrast, no such consolation is available in the case of *Ur-Prior Conditionalization*. If *Evidence Externalism* is true and an agent adopts a regular ur-prior, she can't be sure independently of all empirical evidence that she won't ever violate *Global Factivity* or *Global Negative Introspection*. So, even if she in fact takes the utmost care in designing her learning experiences in the course of her cognitive career, conditionalizing plans won't maximize expected accuracy relative to her ur-prior. Relative to her ur-prior, therefore, it will be rationally impermissible for her to conform to the

---

[33] For criticisms of the anti-luminosity argument, see Weatherson (2004), Berker (2008), Ramachandran (2009), Vogel (2010), and Smithies (2012). For criticisms of the anti-KK argument, see Stalnaker (2015), Greco (2014), and Das & Salow (2018).

doxastic plan recommended by *Ur-Prior Conditionalization*. The main difference between the two cases lies in this. The prior credence function using which an agent assesses the expected accuracy of updating according to *Bayesian Conditionalization* is sensitive to empirical evidence that she might have about her future learning experiences; by contrast, the rational ur-prior using which she assesses the expected accuracy of updating according to *Ur-Prior Conditionalization* isn't sensitive to any empirical evidence that she might gather about her learning experiences.

It is worth pointing out that the results proved above are generalizable along two dimensions. First of all, even if we give up *Credal Veritism* and embrace a conception of epistemic value on which features other than gradational accuracy can contribute to the value of a doxastic state, we may be able to prove versions of *Theorem 1* and *Corollaries 1* and *2*. As long as the measure by which we assess the epistemic value of a credence function remains strictly proper, the results in question will hold. So, it may indeed be the case that *Evidence Externalism* is incompatible with any argument for *Ur-Prior Conditionalization* that presupposes some version of epistemic utility theory.

Second, we can also prove a pragmatic analogue of *Theorem 1*, which says the following: relative to an agent's ur-prior, the expected value of acts recommended by credence functions that conform to meta-conditionalizing plans is strictly greater than the expected value of acts recommended by credence functions that conform to any other doxastic plan.[34] Next, we can show that relative to partitional frames, since meta-conditionalizing plans recommend the same credence functions as conditionalizing plans, the expected value of acts recommended by credence functions that conform to conditionalizing plans is greater than the expected value of acts endorsed by other doxastic plans. This constitutes a pragmatic argument for *Ur-Prior Conditionalization*.[35] Finally, in the same way as we proved *Corollary 2*, we can show that relative to a nonpartitional frame and a regular ur-prior, the expected value of acts recommended by credence functions that conform to meta-conditionalizing plans is strictly greater the expected value of acts recommended by credence functions that conform to conditionalizing plans. Therefore, from a purely pragmatic standpoint, *Ur-Prior Conditionalization* is suboptimal.

What this suggests to me is that there is a general tension between *Evidence Externalism* and a certain class of *consequentialist arguments* for *Ur-Prior Conditionalization*, i.e., arguments that require that rational agents, in choosing their doxastic plans, maximize expected value of one kind or another.

---

[34] We can show this simply by adapting a proof given by Brown (1976).

[35] A standard Diachronic Dutchbook argument isn't straightforwardly available for *Ur-Prior Conditionalization*. The standard diachronic Dutchbook argument for *Bayesian Conditionalization* proceeds in two stages: before the agent learns anything, the bookie offers her a combination of two bets, and then later, depending on the situation, the bookie may make a third wager with the agent. This combination of wagers may not be available in the case of *Ur-Prior Conditionalization*, since there may not be any point of time at which an agent in fact has the ur-prior as her actual credence function. At least, according to many conceptions of the ur-prior, the ur-prior is merely a *hypothetical* probability function that is rationally permissible for the agent to adopt independently of all empirical evidence, but there needn't be any actual stage of inquiry at which a rational agent finds herself without any empirical evidence and therefore adopts that ur-prior as her actual credence function. In the absence of such a Dutchbook argument, the only other obvious pragmatic argument for *Ur-Prior Conditionalization* is the one that I have sketched above. However, if there is such a diachronic Dutchbook argument, it would fail for the reasons that Bronfman (2014) mentions.

§7. Despair.   *Corollary 2* suggests that there isn't any plausible argument which allows us to justify *Ur-Prior Conditionalization* by appealing solely to the epistemic goal of accuracy. To see why, note that the defender of *Ur-Prior Conditionalization* could try to defend her principle against the argument given in the last section, simply by rejecting *Evidence Externalism* or by rejecting the assumption that scoring rules should be strictly proper, or by rejecting *Rational Planning*.

7.1. Rejecting Evidence Externalism.   Consider, first, the option of rejecting *Evidence Externalism*.

The first thing to point out is that the problem for *Ur-Prior Conditionalization* mentioned above may well turn out to be a problem for even the Cartesian or phenomenalist theory of evidence. Distinguish two different kinds of *self-evidentness*.

> *Positive Self-Evidentness*. A centered proposition H is positively self-evident iff, for any centered world $\langle w, \langle i, t \rangle \rangle$, if H is true in $\langle w, \langle i, t \rangle \rangle$, i's evidence in $\langle w, \langle i, t \rangle \rangle$ entails H.
> *Negative Self-Evidentness*. A centered proposition H is negatively self-evident iff, for any centered world $\langle w, \langle i, t \rangle \rangle$, if H is false in $\langle w, \langle i, t \rangle \rangle$, i's evidence in $\langle w, \langle i, t \rangle \rangle$ entails $\sim H$.

Positive self-evidentness is an evidential analogue of what Williamson (2000) calls *luminosity*. Williamson argues that there is no nontrivial condition that is *luminous*, such that if it obtains, we are in a position to know that it obtains. Even though Williamson's argument is cast in terms of *knowledge*, we can easily construct a variant of it for *positive self-evidentness*: the crucial feature that may be taken to be common between knowledge and evidence is that a piece of information can be known or be part of one's evidence only if it is acquired by a safe or reliable mechanism. As I pointed out earlier, many writers have resisted Williamson's anti-luminosity argument (see footnote 33). However, even if we grant that Williamson's anti-luminosity argument fails to show that our phenomenal states aren't positively self-evident, we may still be able to show that our phenomenal states aren't negatively self-evident. For example, someone who expects to be burnt by hot water may touch ice-cold water, and judge that she is undergoing a hot sensation. On a natural Cartesian conception of evidence, our evidence consists all and only of facts about our phenomenal states that we know or are in a position to know by introspection. In this example, plausibly, for the first split-second of her experience, the agent isn't in a position to know that she isn't undergoing a hot sensation. So, the proposition that the agent is undergoing a hot sensation isn't negatively self-evident. Here, the agent's evidence doesn't entail that she is undergoing a hot sensation, but her evidence also doesn't entail that her evidence doesn't entail it. Therefore, if there are propositions about phenomenal states that aren't negatively self-evident, *Global Negative Introspection* can fail even on a Cartesian picture of evidence. If an agent assigns nonzero probability to such failures of *Global Negative Introspection*, the accuracy-based argument for *Ur-Prior Conditionalization* will break down.

However, let's say a defender of the Cartesian conception of evidence is able to show that all propositions about phenomenal states—or at least those that we can know by introspection—are negatively self-evident. What should we say to that person? Recall the skeptical challenge that we raised earlier for the Cartesian picture of evidence: the Cartesian would have to say how we ever come to form justified beliefs about the world in the absence of prior empirical evidence about the reliability of our perceptual and cognitive mechanisms. Some writers, such as Pryor (2000), have responded to this challenge. On

the view that Pryor defends—what he calls *dogmatism*—perceptual experiences provide *prima facie* justification for beliefs about the external world even in the absence of any background evidence about the veridicality of such experiences.

However, such a view is incompatible with *Ur-Prior Conditionalization*. For, if *Ur-Prior Conditionalization* is correct, then, given a body of evidence E_t, an agent can only be rational in believing a proposition H about the external world if $\mu(H|E_t)$ is sufficiently higher than $\mu(\sim H|E_t)$. Now, this can only be the case if $\mu(H \cap E_t)$ is sufficiently higher that $\mu(\sim H \cap E_t)$. But this means that the agent, independently of all empirical evidence, must be considerably more confident in $H \cap E_t$ than in $\sim H \cap E_t$. For example, if E_t is the centered proposition that it appears to me as if I have a hand, and H is the centered proposition that I have a hand, then in order to be justified in believing that I have a hand on the basis of the relevant evidence, I must have antecedent reason to be more confident that I have hands and it appears to me that I have hands than that I don't have hands and it appears to me that I have hands. So, background evidence does matter on this picture.

The only version of the Cartesian story that does not conflict with *Ur-Prior Conditionalization* is the one defended by writers like Wright (2004) and White (2006). Both of these writers deny the assumption that we need empirical evidence in order to be justified in taking our perceptual and cognitive faculties to be reliable. For Wright (2004), we are *entitled* on purely nonevidential grounds (e.g., for pragmatic reasons, for reasons having to do with our projects of inquiry, etc.) to accept the claim that our ordinary methods of belief-formation are reliable. By contrast, for White (2006), we are justified *a priori* in ruling out skeptical possibilities where our perceptual and cognitive faculties mislead us.

Both these views seem to entail that it is rationally permissible for us to have a bias against a class of contingent hypotheses, namely those on which our faculties provide misleading information, independently of all empirical evidence whatsoever. These views might strike us as counterintuitive. Wright's view seems to run contrary to a widely accepted evidentialist approach to epistemic rationality, which requires all agents to proportion their doxastic attitudes to the evidence they possess. White's view, by contrast, licenses a strong form of rationalism, on which we have *a priori* justification for believing certain contingent claims about the world. Therefore, in order to accept the Cartesian picture of evidence, we will indeed have to accept certain controversial commitments about justification and rationality.

Finally, even if we can reconcile ourselves to the Wright-White picture of rationality, it may still be difficult to salvage the accuracy-based argument for *Ur-Prior Conditionalization* for other reasons. The set of centered worlds over which our ur-priors are defined is the set of all *epistemically possible* worlds, i.e., worlds that we can't rule out independently of all empirical evidence. So, if *Partitional Evidence* is to be true, we should be able to rule out, independently of all empirical evidence, the possibility that our methods of gathering evidence—whatever that evidence might be—are fallible, i.e., could malfunction and fail to provide us with evidence, without giving us a warning that this has happened. But it seems that even if our methods of learning about our own phenomenal states are *in fact* infallible, we only learn this on the basis of empirical evidence, e.g., by observing the tight connection between our phenomenal states and our awareness of them. So, it is at least not obvious that we can rule out *a priori* that our methods of learning about our internal world are fallible in the way described above. Hence, even if our ordinary concept of evidence is Cartesian, it is not clear that *Global Negative Introspection* simply falls out of our ordinary concept of evidence; it is not a conceptual truth. And, without *Global Negative Introspection*, the argument for *Ur-prior Conditionalization* won't go through.

Thus, *Ur-Prior Conditionalization* cannot be defended simply by accepting the Cartesian picture of evidence.

7.2. Rejecting strict propriety. Another option for defenders of *Ur-Prior Conditionalization* would be to reject strict propriety, which plays an important role in the proof of *Theorem 1*. For instance, we could require the scoring rule *A* to be *proper*, and not strictly proper.

> *Propriety*. If B_s is the set of all credence functions defined over a set of states S and $A : B_S \times S \to [0, 1]$ is an epistemic scoring rule, A is said to be *proper* iff, for any probability function $p \in B_S$ and any credence function $b \in B_S$ distinct from p, the expected accuracy of p according to p is greater than or equal to the expected accuracy of b according to p, i.e., $Exp_p(p) \geq Exp_p(b)$.

Now, in the same way as we proved *Theorem 1*, we can show

> *Theorem 1**. Suppose $D = \langle W, I, T, \geq, S, E, \mu, B_S, A \rangle$ is a cognitive decision problem where A is a proper epistemic value function. Then, for any doxastic plan $R : S \to B_S$, R maximizes expected accuracy according to μ and A if R is a meta-conditionalizing plan for μ. That is,
>
> > (i) For any two doxastic plans $R : S \to B_S$ and $R^* : S \to B_S$, if both R and R^* are meta-conditionalizing plans for μ, then $Exp_\mu(R) = Exp_\mu(R^*)$.
> > (ii) For any two doxastic plans $R : S \to B_S$ and $R^* : S \to B_S$, if R is a meta-conditionalizing plan for μ but R^* is not, then $Exp_\mu(R) \geq Exp_\mu(R^*)$.[36]

If this correct, then *Ur-Prior Conditionalization* does not constitute a *requirement* of epistemic rationality: since, according to *Rational Planning*, it will always remain rationally permissible for the agent to meta-conditionalize, she can't be required to update her credences by conditionalizing a rationally permissible ur-prior on her total evidence.

The final step might be to find a scoring rule which is improper in the following sense.

> *Impropriety*. If B_s is the set of all credence functions defined over a set of states S and $A : B_S \times S \to [0, 1]$ is an epistemic scoring rule, A is said to be *improper* iff there exists a probability function $p \in B_S$ and a credence function $b \in B_S$ distinct from p such that the expected accuracy of p according to p is less than the expected accuracy of b according to p, i.e., $Exp_p(p) < Exp_p(b)$.

We may indeed be able to justify *Ur-Prior Conditionalization* using an improper rule. But that will have its cost. Propriety, as a constraint on scoring rules, is motivated by a constraint, sometimes called *immodesty*: namely, that a rational agent's credence function shouldn't be suboptimal by her own lights. We can see why this entails that a scoring rule that is appropriate for a rational agent to use must be proper. Otherwise, the expected accuracy of her credence function would be lower than that of another credence function. This would mean that, by her own lights, her own credence function is suboptimal.

---

[36] Since this result is trivial given our proof of *Theorem 1*, I shall not prove it.

This is bad. Joyce explains why.

> If, relative to a person's own credences, some alternative system of beliefs has a lower expected epistemic disutility, then, by her own estimation, that system is preferable from the epistemic perspective. This puts her in an untenable doxastic situation. She has a prima facie epistemic reason, grounded in her beliefs, to think that she should not be relying on those very beliefs. This is a probabilistic version of Moore's paradox. Just as a rational person cannot fully believe "X but I don't believe X," so a person cannot rationally hold a set of credences that require her to estimate that some other set has higher epistemic utility. The modest person is always in this pathological position: her beliefs undermine themselves. (Joyce (2009), p. 277)

The thought is this. When an agent measures accuracy using an improper scoring rule, she may expect credence functions other than the one she rationally adopts as more accurate than her own. In such scenarios, it will indeed be instrumentally rational, from an accuracy-based standpoint, to switch to those other credences. Thus, her credences will be self-undermining. That is why *Impropriety* seems unappealing.

7.3. Rejecting Rational Planning.
The only other option is to reject *Rational Planning*. If we do not want to reject the *accuracy-first* approach to epistemology, we need to modify or replace *Rational Planning* while staying within the limits of that approach. But how could we do this?

According to *Rational Planning*, relative to a cognitive decision problem $D = \langle W, I, T, \geq, S, E, \mu, B_S, A \rangle$, it is rationally permissible for an agent to conform to a doxastic plan $R : S \to B_S$ iff, for any doxastic plan $R^* : S \to B_S$, $Exp_\mu(R) \geq Exp_\mu(R^*)$.

A tempting response might be to say that this notion of *Rational Planning* presupposes that expected value maximization is the correct norm of instrumental rationality. Perhaps, we can save *Ur-Prior Conditionalization* by simply rejecting that presupposition. It is hard to see where this response will lead. In discussions of practical rationality, many theorists reject expected value maximization in light of various counterexamples.[37] So, one place to start would be to see whether any of alternative decision rules that these writers propose could be used to vindicate *Ur-Prior Conditionalization*. It is worth pointing out, however, that most of these decision rules yield expected value maximization as a special case.[38] If conditionalizing plans don't maximize expected accuracy for nonpartitional evidence and regular ur-priors, they won't be optimal in general according to these decision rules either.

A more promising response to the problem might be this. Why does conforming to *Ur-Prior Conditionalization* not maximize expected accuracy in cases where *Partitional Evidence* fails? The reason is that only meta-conditionalizing plans maximize expected accuracy in such cases. But one might think that an agent who lacks perfect access to her own evidence won't be able to *competently execute* meta-conditionalizing plans.[39] This is because, when an agent lacks access to her own evidence, the agent's total evidence won't entail what meta-conditionalizing requires her to do in that scenario. Now, suppose we also say that an agent can only be rationally required to conform to a plan if she can competently execute it. So, meta-conditionalizing cannot be a rational requirement.

---

[37] For a survey, see Briggs (2017).

[38] For a survey, see Starmer (2000).

[39] This idea, though intuitive enough, was first developed by Bronfman (2014).

However, this view also impugns *Ur-Prior Conditionalization*. When *Partitional Evidence* fails due to a failure of *Global Negative* or *Positive Introspection*, then an agent will lack access to her own evidence. Just as the agent's total evidence doesn't entail what meta-conditionalizing involves in that scenario, so also her total evidence won't entail what conditionalizing involves in that scenario. So, if this prevents an agent from competently executing meta-conditionalizing plans, so also should it prevent her from competently executing conditionalizing plans. So, if our aim is to defend *Ur-Prior Conditionalization*, this doesn't seem to be the best strategy.

Of course, I haven't ruled out every possible strategy for rehabilitating *Ur-Prior Conditionalization* by rejecting *Rational Planning*. But, still, I hope to have made clear why there is no easy way of rejecting *Rational Planning* that would help the defender of *Ur-Prior Conditionalization* justify her principle on the basis of considerations about accuracy.

§8. Consequences. In the last section, I argued that the defender of accuracy-first epistemologist cannot straightforwardly justify *Ur-Prior Conditionalization* even if she rejects *Evidence Externalism*, the strict propriety of scoring rules or *Rational Planning*. In response to this argument, a hard-nosed accuracy-first epistemologist might be tempted to reject *Ur-Prior Conditionalization*, and say that an agent is rationally required to adopt meta-conditionalizing plans relative to her ur-prior rather than conditionalizing plans. But this commits her to the following claim.

> *Perfect Access*. If an agent's total evidence is E, she is required by epistemic rationality to be certain that her total evidence is E.

If the accuracy-first epistemologist accepts *Perfect Access* (which she must), she must either reject the possibility of failures of *Global Positive Introspection* and *Global Negative Introspection* or give up the *Principle of Total Evidence*.

To see why, consider a scenario like *Red Wall*, where *Global Negative Introspection* fails. When the agent looks at the white wall lit up with red light, she learns nothing about the color of the wall. But her evidence can't rule out the possibility s_{r2} that the wall before her is red; in fact, that possibility might be highly likely on her total evidence. However, if *Perfect Access* is correct, she must be certain that her evidence doesn't rule out s_{w2} either. From that, she can conclude that she is not in s_{r2} and therefore isn't looking at a red wall. Hence, *Perfect Access* can allow an agent be certain in certain propositions that are extremely unlikely on their evidence. But surely that violates the *Principle of Total Evidence*, which requires the agent to adopt doxastic attitudes that reflect the degrees of support that the relevant propositions get from her total evidence. Thus, the accuracy-first epistemologist faces a dilemma: since she is committed to *Perfect Access*, she must either reject the possibility of introspection failures like *Red Wall* or reject the *Principle of Total Evidence*.

The significance of this dilemma is two-fold. First, some defenders of *Ur-Prior Conditionalization*—such as Williamson (2000) and Hedden (2015)—reject both *Global Positive* and *Negative Introspection*, and yet are committed to the *Principle of Total Evidence*. The dilemma presented above shows that such writers cannot rely on accuracy-based arguments in order to defend *Ur-Prior Conditionalization*.[40] Second, the dilemma posed above also calls into question the accuracy-first approach to epistemology. It shows

---

[40] Still, defenders of *Ur-Prior Conditionalization* might hope to justify *Ur-Prior Conditionalization* by appealing to a different kind of epistemic value. But I think that there is little hope of doing

that the accuracy-first approach to epistemology cannot be reconciled with the *Principle of Total Evidence* unless we accept extremely strong constraints on the structure of evidence, such as *Global Negative Introspection*. As I have pointed out above, *Global Negative Introspection* should be treated with suspicion not only by defenders of *Evidence Externalism*, but also by those who are sympathetic to a Cartesian or phenomenalist conception of evidence. So, I take the dilemma presented above to be bad news for accuracy-first epistemology.[41]

§9. **Appendix: Proofs.** Let us begin with the proof of *Theorem 1*.

Theorem 1. Suppose $D = \langle W, I, T, \geq, S, E, \mu, B_S, A \rangle$ is a cognitive decision problem where A is a strictly proper scoring rule. Then, for any doxastic plan $R : S \to B_S$, R maximizes expected accuracy according to μ and A iff R is a meta-conditionalizing plan for μ
That is,

(i) For any two doxastic plans $R : S \to B_S$ and $R^* : S \to B_S$, if both R and R^* are meta-conditionalizing plans for μ, then $Exp_\mu(R) = Exp_\mu(R^*)$.
(ii) For any two doxastic plans $R : S \to B_S$ and $R^* : S \to B_S$, if R is a meta-conditionalizing plan for μ but R^* is not, then $Exp_\mu(R) > Exp_\mu(R^*)$.

Proof. Let $\Pi = \{X \subseteq S : (\exists s \in S)(E(s) = X)\}$ be the set of all possible bodies of total evidence that one could have in S. Let R_X be the credence function a doxastic plan R outputs in s iff $E(s)$ is X.
The expected accuracy of a doxastic plan R according to μ is

$$
\begin{aligned}
Exp_\mu(R) &= \sum_{s \in S} \mu(s) A(R(s), s) \\
&= \sum_{X \in \Pi} \sum_{s \in [\mathbf{E}=X]} \mu(s) A(R(s), s) \\
&= \sum_{X \in \Pi} \sum_{s \in [\mathbf{E}=X]} \mu(\{s\} \cap [\mathbf{E} = X]) A(R(s), s) \\
&= \sum_{X \in \Pi, \mu([\mathbf{E}=X])>0} \mu([\mathbf{E} = X]) \sum_{s \in [\mathbf{E}=X])} \mu(s|[\mathbf{E} = X]) A(R(s), s) \\
&= \sum_{X \in \Pi, \mu([\mathbf{E}=X])>0} \mu([\mathbf{E} = X]) \sum_{s \in S} \mu(s|[\mathbf{E} = X]) A(R(s), s).
\end{aligned}
$$

so. Note that *Corollary 2* does not in any way depend on the epistemic value function A's being a function that measures accuracy. As long as A is strictly proper or proper in the sense specified above, the status of *Ur-Prior Conditionalization* as a requirement of epistemic rationality cannot be justified. And, as we have already seen, impropriety leads to instability. So, I don't think epistemic utility theory can be used for justifying *Ur-Prior Conditionalization*.

[41] I am grateful to Daniel Kokotajlo, Barry Maguire, and members of the UNC Chapel Hill Formal Epistemology Reading Group for comments on this article.

Now, for any two doxastic plans $R : S \rightarrow B_S$ and $R^* : S \rightarrow B_S$, if both R and R^* are meta-conditionalizing plans for μ, then, for any evidence-proposition $X \in \Pi$ such that $\mu([\mathbf{E} = X]) > 0$, R and R^* will recommend the same credence function everywhere. So, $Exp_\mu(R) = Exp_\mu(R^*)$. This establishes (i).

However, if R is a meta-conditionalizing plan for μ, but R^* isn't, then there will be one evidence-proposition $X \in \Pi$ such that $\mu([\mathbf{E} = X]) > 0$, but R and R^* recommend different credences. But since A is strictly proper, $\sum_{s \in S} \mu(s|[\mathbf{E} = X])A(R_X, s) > \sum_{s \in S} \mu(s|[\mathbf{E} = X])A(R_X^*, s)$. As a result, $Exp_\mu(R) > Exp_\mu(R^*)$. This establishes (ii).

Q. E. D.

To derive *Corollary 1* from *Theorem 1*, we need the following lemma.

Lemma. Suppose $\langle W, I, T, \geq, S, E, \mu \rangle$ is a self-locating frame. Let $\Pi = \{X \subseteq S : (\exists s \in S)(E(s) = X)\}$ be the set of all possible bodies of total evidence that one could have in S. Then, $\langle W, I, T, \geq, S, E, \mu \rangle$ is partitional iff, for any centered proposition $X \in \Pi$, $[\mathbf{E} = X] = X$.

Proof. We need to prove two conditionals:

Conditional 1. If $\langle W, I, T, \geq, S, E, \mu \rangle$ is partitional, then, for any centered proposition $X \in \Pi$, $[\mathbf{E} = X] = X$.

Conditional 2. If $\langle W, I, T, \geq, S, E, \mu \rangle$ is not partitional, then there exists a centered proposition $X \in \Pi$ such that $[\mathbf{E} = X] \neq X$.

First, we prove *Conditional 1*. Suppose $\langle W, I, T, \geq, S, E, \mu \rangle$ is partitional. Then,

1. For any $s \in S$, if $s \in [\mathbf{E} = X]$, then $E(s) = X$. By reflexivity, $s \in E(s)$, so $s \in X$. Therefore, $[\mathbf{E} = X] \subseteq X$.

2. For any $s \in X$, if $s \notin [\mathbf{E} = X]$, then $E(s) \neq X$. Then, either (i) there exists a state $s^* \in E(s)$ such that $s^* \notin X$, or (ii) there exists a state $s^* \in X$ such that $s^* \notin E(s)$.

(a) (i) cannot be correct. Since $X \in \Pi$, there exists a state s^{**} such that $E(s^{**}) = X$. If (i) is correct, we have a scenario where $s \in E(s^{**}) = X$, and $s^* \in E(s)$, but $s^* \notin E(s^{**}) = X$. This violates transitivity.

(b) (ii) cannot be correct either. Since $X \in \Pi$, there exists a state s^{**} such that $E(s^{**}) = X$. If (ii) is correct, we have a scenario where $s \in E(s^{**}) = X$, and $s^* \in E(s^{**})$, but $s^* \notin E(s) = X$. This violates euclideanness.

So, for any $s \in X$, if $s \in X$, then $s \in [\mathbf{E} = X]$. Therefore, $X \subseteq [\mathbf{E} = X]$.

Therefore, for any centered proposition $X \in \Pi$, $[\mathbf{E} = X] = X$. Thus, *Conditional 1* is proved.

Next, we prove *Conditional 2*. Suppose $\langle W, I, T, \geq, S, E, \mu \rangle$ is not partitional. There are three possibilities: either it is nonreflexive, or nontransitive, or noneuclidean.

1. If $\langle W, I, T, \geq, S, E, \mu \rangle$ is not reflexive, then there exists a centered world $s \in S$ and a centered proposition X such that $E(s) = X$, but $s \notin X$. In that case, $[\mathbf{E} = E(s)] \neq E(s)$.

2. If $\langle W, I, T, \geq, S, E, \mu \rangle$ is not transitive, then there are two centered worlds $s, s^* \in S$ such that $s^* \in E(s)$, but $E(s^*)$ is not a subset of $E(s)$. So, $s^* \in E(s)$, but $s^* \notin [\mathbf{E} = E(s)]$. In that case, $[\mathbf{E} = E(s)] \neq E(s)$.

3. If $\langle W, I, T, \geq, S, E, \mu \rangle$ is not euclidean, then there are three centered worlds $s, s^*, s^{**} \in S$ such that $s^* \in E(s)$ and $s^{**} \in E(s)$, but $s^{**} \notin E(s^*)$. If the frame is reflexive, this means that $E(s^{**}) \neq E(s^*)$. This implies that $[\mathbf{E} = E(s)] \neq E(s)$. For, if $[\mathbf{E} = E(s)]$ were equal to $E(s)$, then it would be the case that $E(s^{**}) = E(s^*) = E(s)$.

Thus, *Conditional 2* is proved.

Q. E. D.

Now, let us focus on *Corollary 1*.

> *Corollary 1.* Suppose $D = \langle W, I, T, \geq, S, E, \mu, B_S, A \rangle$ is a cognitive decision problem where $\langle W, I, T, \geq, S, E, \mu \rangle$ is a partitional self-locating frame and A is a strictly proper scoring rule. Then, for any doxastic plan $R : S \to B_S$, R maximizes expected accuracy according to μ and A iff R is a conditionalizing plan for μ.
> That is,
>
> > (i) For any two doxastic plans $R : S \to B_S$ and $R^* : S \to B_S$, if both R and R^* are conditionalizing plans for μ, then $Exp_\mu(R) = Exp_\mu(R^*)$.
> >
> > (ii) For any two doxastic plans $R : S \to B_S$ and $R^* : S \to B_S$, if R is a conditionalizing plan for μ but R^* is not, then $Exp_\mu(R) > Exp_\mu(R^*)$.
>
> *Proof.* $\langle W, I, T, \geq, S, E, \mu \rangle$ is a partitional self-locating frame Note that, in the proof of *Theorem 1*, we got
>
> $$Exp_\mu(R) = \sum_{X \in \Pi, \mu([\mathbf{E}=X]) > 0} \mu([\mathbf{E} = X]) \sum_{s \in S} \mu(s | [\mathbf{E} = X]) A(R_X, s).$$
>
> But, since the frame is partitional, by *Lemma*, for any centered proposition $X \in \Pi$, $[\mathbf{E} = X] = X$. Hence, for any μ and any $X \in \Pi$, $\mu(. | [\mathbf{E} = X]) = \mu(. | X)$, provided $\mu(X) > 0$. So, every conditionalizing plan will be a meta-conditionalizing plan. Therefore, the following will be true.
>
> > (i) For any two doxastic plans $R : S \to B_S$ and $R^* : S \to B_S$, if both R and R^* are conditionalizing plans for μ, then $Exp_\mu(R) = Exp_\mu(R^*)$.

(ii) For any two doxastic plans $R : S \rightarrow B_S$ and $R^* : S \rightarrow B_S$, if R is a conditionalizing plan for μ but R^* is not, then $Exp_\mu(R) > Exp_\mu(R^*)$.

Q. E. D.

Next, we prove *Corollary 2*.

Corollary 2. Suppose $D = \langle W, I, T, \geq, S, E, \mu, B_S, A \rangle$ is a cognitive decision problem where A is a strictly proper scoring rule. Then, the following are inconsistent.

(i) For any two doxastic plans $R : S \rightarrow B_S$ and $R^* : S \rightarrow B_S$, if R is a conditionalizing plan relative to μ, $Exp_\mu(R) \geq Exp_\mu(R^*)$.
(ii) μ is a *regular* probability function, i.e., for any $s \in S$, $\mu(s) > 0$.
(iii) $\langle W, I, T, \geq, S, E, \mu \rangle$ is a nonpartitional self-locating frame.

Proof. Suppose μ is a *regular* probability function, and $\langle W, I, T, \geq, S, E, \mu \rangle$ is a nonpartitional self-locating frame.

If $\langle W, I, T, \geq, S, E, \mu \rangle$ is nonpartitional, then, by *Lemma*, there exists a proposition $X \in \Pi$ such that $[\mathbf{E} = X] \neq X$. That is, there exists a world s such that either (i) $s \in X$ but $s \notin [\mathbf{E} = X]$ or (ii) $s \notin X$ but $s \in [\mathbf{E} = X]$.

However, since μ is regular, $\mu(s) > 0$. So, either $\mu(s|[\mathbf{E} = X]) = 0$ but $\mu(s|X) > 0$, or $\mu(s|[\mathbf{E} = X]) > 0$ but $\mu(s|X) = 0$. Hence, $\mu(.|[\mathbf{E} = X]) \neq \mu(.|X)$. Therefore, no conditionalizing plan for μ will be a meta-conditionalizing plan for μ. By *Theorem 1*, it follows that if R is a meta-conditionalizing plan and R^* is a conditionalizing plan, then $Exp_\mu(R) > Exp_\mu(R^*)$.

Q.E.D.

BIBLIOGRAPHY

Adler, J. (2002). *Belief's Own Ethics*. Cambridge, MA: MIT Press.

Adler, J. E. (1989). Epistemics and the total evidence requirement. *Philosophia*, **19**(2–3), 227–243.

Alston, W. P. (1980). Level-confusions in epistemology. *Midwest Studies in Philosophy*, **5**(1), 135–150.

Antony, L. (2004). A naturalized approach to the a priori. *Philosophical Issues*, **14**(1), 1–17.

Berker, S. (2008). Luminosity regained. *Philosophers' Imprint*, **8**, 1–22.

Bradley, D. (2011). Self-location is no problem for conditionalization. *Synthese*, **182**(3), 393–411.

Briggs, R. A. (2017). Normative theories of rational choice: Expected utility. In Zalta, E. N., editor. *The Stanford Encyclopedia of Philosophy*, Spring 2017 Edition. Available at https://plato.stanford.edu/entries/rationality-normative-utility/.

Briggs, R. A. & Pettigrew, R. (2016). An accuracy dominance argument for conditionalization. Unpublished manuscript.

Bronfman, A. (2014). Conditionalization and not knowing that one knows. *Erkenntnis*, **79**(4), 871–892.

Brown, P. M. (1976). Conditionalization and expected utility. *Philosophy of Science*, **43**(3), 415–419.

Byrne, A. (2013). Perception and evidence. *Philosophical Studies*, **170**(2), 101–113.

Carnap, R. (1962). *Logical Foundations of Probability*. Chicago: University of Chicago Press.

Conee, E. & Feldman, R. (2004). *Evidentialism*. Oxford: Oxford University Press.

Conee, E. & Feldman, R. (2008). Evidence. In Smith, Q., editor. *Epistemology: New Essays*. Oxford: Oxford University Press, pp. 83–104.

Das, N. & Salow, B. (2018). Transparency and the KK principle. *Noûs*, **52**(1), 3–23.

Davidson, D. (2004). Deception and division. In Davidson. *Problems of Rationality*. Oxford: Oxford University Press, pp. 198–212.

Dretske, F. (2004). Externalism and modest contextualism. *Erkenntnis*, **61**(2–3), 173–186.

Easwaran, K. (2013). Expected accuracy supports conditionalization—and conglomerability and reflection. *Philosophy of Science*, **80**(1), 119–142.

Easwaran, K. (2014). Regularity and hyperreal credences. *Philosophical Review*, **123**(1), 1–41.

Elga, A. (2000). Self-locating belief and the sleeping beauty problem. *Analysis*, **60**(266), 143–147.

Fantl, J. & McGrath, M. (2009). *Knowledge in an Uncertain World*. Oxford: Oxford University Press.

Feldman, R. & Conee, E. (1985). Evidentialism. *Philosophical Studies*, **48**(1), 15–34.

Gibbard, A. (2003). *Thinking How to Live*. Cambridge, MA: Harvard University Press.

Goldman, A. (2009). Williamson on knowledge and evidence. In Greenough, P., Pritchard, D., and Williamson, T., editors. *Williamson on Knowledge*. Oxford: Oxford University Press, pp. 73–91.

Greaves, H. & Wallace, D. (2006). Justifying conditionalization: Conditionalization maximizes expected epistemic utility. *Mind*, **115**(459), 607–632.

Greco, D. (2014). Could KK be OK? *Journal of Philosophy*, **111**(4),169–197.

Greco, D. & Hedden, B. (2016). Uniqueness and metaepistemology. *Journal of Philosophy*, **113**(8), 365–395.

Hájek, A. (2003). What conditional probability could not be. *Synthese*, **137**(3), 273–323.

Halpern, J. (2004). Sleeping beauty reconsidered: Conditioning and reflection in asynchronous systems. In Gendler, T. S. and Hawthorne, J., editors. *Oxford Studies in Epistemology*, Vol. 1. Oxford: Oxford University Press, pp. 111–142.

Hedden, B. (2015). *Reasons Without Persons: Rationality, Identity, and Time*. Oxford: Oxford University Press.

Hempel, C. (1965). *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: The Free Press.

Hintikka, J. (1962). *Knowledge and Belief*. Ithaca: Cornell University Press.

Horowitz, S. (2014). Immoderately rational. *Philosophical Studies*, **167**(1), 41–56.

Joyce, J. (2009). Accuracy and coherence: Prospects for an alethic epistemology of partial belief. In Huber, F. and Schmidt-Petri, C., editors. *Degrees of Belief*. Dordrecht: Springer, pp. 263–297.

Joyce, J. M. (1998). A nonpragmatic vindication of probabilism. *Philosophy of Science*, **65**(4), 575–603.

Joyce, J. (2004). Williamson on evidence and knowledge. *Philosophical Books*, **45**(4), 296–305.

Kelly, T. (2008). Evidence. In Zalta, E. N., editor. *Stanford Encyclopedia of Philosophy*, Winter 2016 Edition. Available at https://plato.stanford.edu/archives/win2016/entries/evidence.

Kelly, T. (2008). Evidence: Fundamental concepts and the phenomenal conception. *Philosophy Compass*, **3**(5), 933–955.

Konek, J. (forthcoming). Epistemic conservativity and imprecise credence. *Philosophy and Phenomenological Research*, to appear.

Lasonen-Aarnio, M. (2013). The dogmatism puzzle. *Australasian Journal of Philosophy*, **92**(3), 1–16.

Leite, A. (2013). But that's not evidence; it's not even true. *Philosophical Quarterly*, **63**(250), 81–104.

Lewis, D. (1971). Immodest inductive methods. *Philosophy of Science*, **38**(1), 54–63.

Lewis, D. (1979). Attitudes *de dicto* and *de se*. *Philosophical Review*, **88**(4), 513–543.

Littlejohn, C. (2012). *Justification and the Truth-Connection*. Cambridge: Cambridge University Press.

Mayo-Wilson, C. & Wheeler, G. (2016). Scoring imprecise credences: A mildly immodest proposal. *Philosophy and Phenomenological Research*, **92**(1), 55–78.

McDowell, J. (1982). Criteria, defeasibility, and knowledge. *Proceedings of the British Academy*, **68**, 455–479.

McDowell, J. (1995). Knowledge and the internal. *Philosophy and Phenomenological Research*, **55**(4), 877–893.

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

McGee, V. (1994). Learning the impossible. In Eells, E. and Skyrms, B., editors. *Probability and Conditionals: Belief Revision and Rational Decision*. Cambridge: Cambridge University Press, pp. 179–199.

Meacham, C. J. G. (2008). Sleeping beauty and the dynamics of de se beliefs. *Philosophical Studies*, **138**(2), 245–269.

Meacham, C. J. G. (2016). Ur-priors, conditionalization, and ur-prior conditionalization. *Ergo*, **3**, 444–492.

Moss, S. (2012). Updating as communication. *Philosophy and Phenomenological Research*, **85**(2), 225–248.

Neta, R. & Pritchard, D. (2007). Mcdowell and the new evil genius. *Philosophy and Phenomenological Research*, **74**(2), 381–396.

Oddie, G. (1997). Conditionalization, cogency, and cognitive value. *British Journal for the Philosophy of Science*, **48**(4), 533–541.

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

Pryor, J. (2000). The skeptic and the dogmatist. *Noûs*, **34**(4), 517–549.

Quine, W. V. (1969). Propositional objects. In Quine. *Ontological Relativity and Other Essays*. New York: Columbia University Press, pp. 139–160.

Ramachandran, M. (2009). Anti-luminosity: Four unsuccessful strategies. *Australasian Journal of Philosophy*, **87**(4), 659–673.

Rinard, S. (2015). No exception for belief. *Philosophy and Phenomenological Research*, **91**(2), 121–143.

Salmon, W. C. (1967). *The Foundations of Scientific Inference*. Pittsburgh: University of Pittsburgh Press.

Schafer, K. (2014). Doxastic planning and epistemic internalism. *Synthese*, **191**(12), 2571–2591.

Schoenfield, M. (2013). Permission to believe: Why permissivism is true and what it tells us about irrelevant influences on belief. *Noûs*, **47**(1), 193–218.

Schoenfield, M. (2017). Conditionalization does not maximize expected accuracy. *Mind*, **126**(504), 1155–1187.

Schoenfield, M. (forthcoming). Permissivism and the value of rationality. *Philosophy and Phenomenological Research*, to appear.

Schultheis, G. (forthcoming). Living on the edge: Against epistemic permissivism. *Mind*, to appear.

Schwarz, W. (2012). Changing minds in a changing world. *Philosophical Studies*, **159**(2), 219–239.

Silins, N. (2005). Deception and evidence. *Philosophical Perspectives*, **19**(1), 375–404.

Skyrms, B. (1983). Three ways to give a probability assignment a memory. In Earman, J. editor. *Testing Scientific Theories*. Minneapolis: University of Minnesota Press, pp. 157–161.

Smithies, D. (2012). Mentalism and epistemic transparency. *Australasian Journal of Philosophy*, **90**(4), 723–741.

Sober, E. (1975). *Simplicity*. Oxford: Clarendon Press.

Sober, E. (2009). Absence of evidence and evidence of absence: Evidential transitivity in connection with fossils, fishing, fine-tuning, and firing squads. *Philosophical Studies*, **143**(1), 63–90.

Stalnaker, R. (2006). On logics of knowledge and belief. *Philosophical Studies*, **128**(1), 169–199.

Stalnaker, R. (2008). *Our Knowledge of the Internal World*. Oxford: Oxford University Press.

Stalnaker, R. (2009). On Hawthorne and Magidor on assertion, context, and epistemic accessibility. *Mind*, **118**(470), 399–409.

Stalnaker, R. (2015). Luminosity and the KK thesis. In Goldberg, S. C., editor. *Externalism, Self-Knowledge, and Skepticism*, Vol. 1. Cambridge: Cambridge University Press, pp. 167–196.

Starmer, C. (2000). Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of Economic Literature*, **38**(2), 332–382.

Talbott, W. J. (1991). Two principles of Bayesian epistemology. *Philosophical Studies*, **62**(2), 135–150.

Teller, P. (1973). Conditionalization and observation. *Synthese*, **26**(2), 218–258.

Titelbaum, M. G. (2014). *Quitting Certainties: A Bayesian Framework Modeling Degrees of Belief*. Oxford: Oxford University Press.

Titelbaum, M. G. & Kopec, M. (forthcoming). When rational reasoners reason differently. In Balcerak-Jackson, M. and Balcerak-Jackson, B., editors. *Reasoning: Essays on Theoretical and Practical Thinking*. Oxford: Oxford University Press.

Van Fraassen, B. C. (1989). *Laws and Symmetry*. Oxford: Oxford University Press.

van Fraassen, B. C. (1999). Conditionalization, a new argument for. *Topoi*, **18**(2), 93–96.

Vogel, J. (2010). Luminosity and indiscriminability. *Philosophical Perspectives*, **24**(1), 547–572.

Weatherson, B. (2004). Luminous margins. *Australasian Journal of Philosophy*, **82**(3), 373–383.

White, R. (2005). Epistemic permissiveness. *Philosophical Perspectives*, **19**(1), 445–459.

White, R. (2006). Problems for Dogmatism. *Philosophical Studies*, **131**(3), 525–557.

Williams, M. (1991). *Unnatural Doubts: Epistemological Realism and the Basis of Scepticism*. B. Blackwell.

Williams, P. M. (1980). Bayesian conditionalisation and the principle of minimum information. *British Journal for the Philosophy of Science*, **31**(2), 131–144.

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

Wright, C. (2004). Warrant for nothing (and foundations for free). *Aristotelian Society Supplementary Volume*, **78**(1), 167–212.

HUMANITIES DIVISION (PHILOSOPHY)
NEW YORK UNIVERSITY SHANGHAI
1555 CENTURY AVENUE
PUDONG NEW AREA, SHANGHAI 200122
CHINA
E-mail: ndas@nyu.edu