

JUST AS PLANNED: BAYESIANISM, EXTERNALISM, AND PLAN COHERENCE

Pablo Zendejas Medina

New York University

*This work is licensed under a Creative Commons
Attribution-NonCommercial-NoDerivatives 3.0 License
<www.philosophersimprint.org/023028/>*

Abstract

Two of the most influential arguments for Bayesian updating (“Conditionalization”) — Hilary Greaves’ and David Wallace’s *Accuracy Argument* and David Lewis’ *Diachronic Dutch Book Argument* — seem to impose a strong and surprising limitation on rational uncertainty: that one can never be rationally uncertain of what one’s evidence is. Many philosophers (“externalists”) reject that claim, and now seem to face a difficult choice: either to endorse the arguments and give up Externalism, or to reject the arguments and lose some of the best justifications of Conditionalization. I argue that the key to resolving this conflict lies in recognizing that both arguments are *plan-based*, in that they argue for Conditionalization by first arguing that one ought to *plan* to conditionalize. With this in view, I argue that the conflict with Externalism only arises if one misconceives the requirement to carry out a plan made at an earlier time. These arguments should therefore not persuade us to reject Externalism. Furthermore, rethinking the nature of this requirement allows us to give two new arguments for Conditionalization that don’t rule out rational uncertainty about one’s evidence and that can thus serve as common ground in the debate between externalists and their opponents.

1. Introduction

The aim of this paper is to reconcile two important ideas in epistemology: the Bayesian view that rational beliefs evolve by conditionalization on evidence, and the externalist view that one can rationally be uncertain about what one’s evidence is, and thus also about what one should rationally believe. More precisely, the first view is:

CONDITIONALIZATION Upon learning that E , a rational agent will conditionalize their prior credences on E , so that, for any

proposition p ,

$$C_E(p) = C(p|E) =_{\text{df}} \frac{C(p \wedge E)}{C(E)},$$

if defined, where C and C_E are the agent's credence functions before and after learning that E , respectively.

This claim is widely assumed and applied not only in philosophy, but in many fields that model belief probabilistically, such as statistics, economics, and computer science. There are two parts to this principle. First, it says that one's beliefs should evolve by conditionalizing on *some* proposition. To conditionalize on a proposition E is to revise one's beliefs so that one's posterior credence in any claim p is directly proportional to one's prior credence in the conjunction of p and E , renormalized so that E is now assigned a credence of one. The second part of CONDITIONALIZATION says that the proposition on which one should conditionalize is one's *evidence*. This second part is not superfluous: "evidence" is not just what we call whatever the agent happens to conditionalize on. Rather, as I'll understand it, what evidence the agent has — what they have *learnt* — is fixed outside of the Bayesian model of which CONDITIONALIZATION forms part.¹ In other words, CONDITIONALIZATION says nothing about the conditions for learning different propositions, but is only an answer to the question: "How should my beliefs change, *given* the evidence I have learnt?"

Two of the most influential arguments for CONDITIONALIZATION, which I will call the *standard arguments*, are David Lewis' (1999) *Diachronic Dutch Book Argument* and Hilary Greaves & David Wallace's (2006) *Accuracy Argument*. Both aim to establish CONDITIONALIZATION by first showing that one should *plan* to conditionalize. My aim is to reconcile this style of argument with:

EXTERNALISM (informal) It is possible for a rational agent to learn that E , without thereby becoming certain that they learnt that E .²

Epistemic externalists typically think that the factors that ground the rationality of our beliefs can sometimes fail to be accessible to us. Since, as I'll assume, one's evidence determines what one can rationally believe, rational uncertainty about the former is one of the main sources of rational uncertainty about the latter.

To be clear: CONDITIONALIZATION itself is *consistent* with EXTERNALISM. A conditionalizer can be uncertain about what they have learnt if their evidence is *opaque* in the sense that it is not equivalent to the claim that it is their evidence. Indeed, externalists typically hold their view because they believe that it is possible to have opaque evidence and that one should conditionalize on it (more on this in §§3.2). Rather, the conflict arises because the standard arguments are restricted to situations where one cannot learn opaque evidence. Moreover, several recent papers (Hild, 1998; Schoenfield, 2017) have claimed that if we generalize these arguments to situations where opaque evidence is possible, they no longer support CONDITIONALIZATION, but rather an alternative update rule, called AUTO-EPISTEMIC (A-E) CONDITIONALIZATION. This rule says that if E is one's evidence, one should conditionalize not on E , but on the claim that E is one's evidence, and is thus, as I'll explain, inconsistent with EXTERNALISM.

I will argue that the key to resolving this conflict lies in the widely recognized fact that both arguments are *plan-based* in that they try to establish CONDITIONALIZATION by showing that a rational agent will *plan* to conditionalize on future evidence. This claim only entails CONDITIONALIZATION if combined with a *plan coherence* principle, which says that a rational agent will implement that plan after learning the evidence. However, there are two natural ways of interpreting the claim that a rational agent will implement their antecedent plan, which

1. Schoenfield (2017) calls this "exogenous evidence".

2. I state this claim formally in §3.

differ in the kinds of plan that they require the agent to implement. These are equivalent when the agent cannot learn opaque evidence, but come apart in other cases. While recent literature has generalized the two arguments by implicitly adopting one of these two principles, I argue that it should be rejected in favour of the other. This allows the externalist to reject a premise in the recent arguments for A-E CONDITIONALIZATION and thus against EXTERNALISM.

Moreover, as I show, my preferred plan coherence principle forms the basis for two new arguments for CONDITIONALIZATION. The first is accuracy-theoretic, and builds on Greaves & Wallace's argument. The second argument involves practical planning, rather than planning for what to believe. As such, the second argument is also of interest independently of the question of EXTERNALISM, since it demonstrates a way of arguing for CONDITIONALIZATION from rather minimal commitments about practical rationality and its relationship to epistemic rationality. Most importantly, however, both arguments show that one should conditionalize regardless of whether one's evidence is opaque or not, making them consistent with EXTERNALISM without presupposing it.

2. Framework

We are interested in a *rational* agent who undergoes a learning experience, and in the relationship between their beliefs before and after that experience. I'll call such a situation an *experiment*. We'll focus on the *strongest* proposition that the agent learns in this experiment — their *total incremental evidence*. For simplicity, I will refer to it as the proposition that they have *learnt*, or as their *evidence*, but keep in mind that this is distinct from their total *background* evidence, which includes what the agent already knows going into the experiment. We'll use " E_1, E_2, \dots " to stand for the different propositions that the agent may learn in the experiment. For any E , moreover, we'll let $\mathbb{L}E$ stand for the

proposition that says that the agent learns that E .³

Our agent begins the experiment with a *rational* credence function C (their "prior"), which is defined over the set of propositions. After undergoing the experiment, they then adopt a new rational credence function C_E (their "posterior") where E is their evidence. Formally, propositions are subsets of a set \mathcal{W} of doxastic possibilities ("worlds"), which we assume is finite for mathematical simplicity. Finally, we'll make two other simplifying, but dispensable, assumptions: that the agent assigns positive credence to all worlds in \mathcal{W} before learning, and that they cannot learn an inconsistent proposition. This ensures that $C(\cdot|E_i)$ is defined for all E_i .⁴

3. Externalism and Evidential Opacity

3.1 Externalism

The externalist about rationality thinks that rational uncertainty about what's rational for one to believe is as possible as rational uncertainty about everyday empirical matters, such as the weather or the outcome of a game. This paper focuses on a specific source of uncertainty about rationality — uncertainty about what one's evidence is. Thus restricted, the externalist thinks that there are possible experiments where one could learn evidence that would leave one rationally uncertain of what one's evidence is. More precisely:

EXTERNALISM (formal) There are possible experiments where the agent can learn some evidence E_i such that $C_{E_i}(\mathbb{L}E_i) < 1$.⁵

Our question is not whether EXTERNALISM is true, but whether the stan-

3. Formally, we can define $\mathbb{L}E$ by letting \mathcal{L} be a function from worlds to propositions, which assigns, to every world, the proposition learnt at that world. Then we let $\mathbb{L}E$ be $\{w \in \mathcal{W} | \mathcal{L}(w) = E\}$.

4. If we did not make this assumption, the conditionalized credence function $C(\cdot|E_i)$ wouldn't be guaranteed to be defined for all E_i , so that there wouldn't always be a unique plan to conditionalize. This would force us to think in terms of *classes* of such plans, unnecessarily complicating the discussion.

5. Since the set of worlds is finite, I understand uncertainty as having a credence below one.

standard Bayesian strategies for justifying update rules can be reconciled with it.

3.2 Evidential Opacity

As is common in Bayesian epistemology, we are treating the agent's evidence as exogenous to the model. We can give it a functional characterization: their evidence is the strongest proposition that their experiences in the experiment enable them to take for granted in their reasoning. But we have not said, and will not say, anything about what it takes for a proposition to have this status. However, externalists typically hold their view in part because they deny a strong, but natural, assumption about evidence. To state it, we'll need the following definition:

Transparent/Opaque Evidence (definition): E is *transparent* if the agent is certain in advance that they will learn that E just in case E is true, i.e. if $C(E \leftrightarrow \mathbb{L}E) = 1$.⁶ If E is not transparent, we say that it is *opaque*.

In practice, when discussing opaque evidence, we'll consider only cases where the agent is uncertain whether they will learn some proposition if it is true (i.e. $C(E \rightarrow \mathbb{L}E) < 1$), not cases where the agent is uncertain whether some proposition is true if they will learn it (i.e. $C(\mathbb{L}E \rightarrow E) < 1$). In this paper, I remain neutral on the possibility of the latter kind of case, and nothing that I say precludes its possibility.

Now, philosophers who accept EXTERNALISM typically do so because they deny that evidence always has to be opaque. That is, they typically deny:

EVIDENTIAL TRANSPARENCY In every possible experiment, every potential body of evidence is transparent.

Intuitively, there is a direct connection between EXTERNALISM and EVIDENTIAL TRANSPARENCY: if you can only ever learn transparent evidence,

6. I use " \rightarrow " and " \leftrightarrow " for the material conditional and biconditional, respectively.

then your evidence will always tell you what your evidence is, so that it would seem irrational to be uncertain of what it is. If, on the other hand, you can have opaque evidence of the kind that leaves open what your evidence is, the right response would seem to be uncertainty about your evidence.

To better understand EVIDENTIAL TRANSPARENCY, and why one might deny it, consider one of the externalist's favourite cases — a sceptical scenario:⁷

Here's a Hand: You're about to open your eyes, and know that you will either have an experience as of a hand (*EXP*) or not. You know that, if you don't have the hand-experience, then you don't have a hand. However, if you do have the hand-experience, you may in fact have a hand (*HAND*), or it could be that you're being deceived by a malicious demon, and there is no hand (\neg *HAND*).

According to the standard externalist treatment of sceptical scenarios your evidential situation is as follows: if you veridically experience the hand — call this the *Good Case* — you will learn that you have a hand, i.e. that you're in the *Good Case*. On the other hand, if you're deceived by the demon — call this the *Bad Case* — you will not learn that you have a hand, but merely that you have an experience as of a hand (Williamson, 2000, ch. 8). What evidence you have is thus determined by facts that are external to you. An externalist may think this for different reasons, but in Williamson's case, this is because he thinks that your evidence is what you come to *know* by looking, which itself depends on whether there is in fact a hand in front of you.

7. For an externalist, sceptical scenarios are counterexamples to EVIDENTIAL TRANSPARENCY because they are failures of *Negative Access* — the claim that if one's evidence doesn't entail some proposition, then it entails that it doesn't entail that proposition. Many externalists also think that there are failures of *Positive Access* — the claim that if one's evidence entails some proposition, it entails that it entails it — such as Williamson's (2014) *Unmarked Clock* case, though this is more controversial.

function — or a *practical act*.¹¹

We can then define two kinds of plans: first, a *conditional plan* is a pair of a condition, which is a proposition, and an act (whether doxastic or practical). So for example, the plan to bring an umbrella if it rains is $\langle \text{Rain}, \text{bring-umbrella} \rangle$, where *Rain* is the proposition that it rains and *bring-umbrella* is the act of bringing an umbrella. Likewise, the plan to conditionalize on E if one learns that E is $\langle \mathbb{I}E, C(\cdot|E) \rangle$. Second, an *unconditional plan* is a set of conditional plans whose conditions partition logical space (\mathcal{W}). So, for example, $\{ \langle \text{Rain}, \text{bring-umbrella} \rangle, \langle \neg \text{Rain}, \text{bring-sunglasses} \rangle \}$ is an unconditional plan, which involves bringing an umbrella if it rains but sunglasses if it doesn't. In the doxastic case, $\text{PCond} = \{ \langle \mathbb{I}E_1, C(\cdot|E_1) \rangle, \langle \mathbb{I}E_2, C(\cdot|E_2) \rangle, \dots \}$ is the plan to conditionalize on one's evidence no matter what one learns, i.e. to conditionalize on E_1 if one learns that E_1 , to conditionalize on E_2 if one learns that E_2 , etc.

4.1.2 Plan Evaluation

We'll assume that the agent has preferences between acts, conditional plans, and unconditional plans, and will say that a plan is *best* in a class if it is strictly preferred to all others in that class. As usual, we'll use " \succeq " to describe the agent's rational (weak) preference relation before learning, as well as " \succ " for strict preference, and " \sim " for indifference.¹²

When is one unconditional plan better than another? The arguments we'll look at answer this question differently, but a common assumption, which most of the arguments below will use is that one should evaluate unconditional plans by their *expected value*. To calculate the expected value of a plan, one takes the weighted sum of the value of the plan at every world, where the value of the plan at a world is just the value, at that world, of the act recommended by the plan at that world. The

weight given to a world is just its credence.¹³ The value function will for the moment be left undefined, since its interpretation differs between arguments.

4.2 Which Plans Trigger Rational Requirements?

We want to argue for *CONDITIONALIZATION* by showing that the plan to conditionalize is the best plan, and that, since one ought to "stick to the plan" after learning, one should conditionalize. Any such argument faces a problem, however: the plan to conditionalize is not best if we compare it to *all* other plans. The standard example is the *omniscience plan*, which says, for every world w , to assign credence one to all truths and credence zero to all falsehoods at w (Greaves & Wallace, 2006, p. 612). On natural ways of evaluating plans, this is the best plan there is, since it will make the agent's beliefs both as accurate (as close to the truth) as can be, and as good a guide to action as one could ask for, since one can never lose by betting on the truth and only on the truth. And yet there is no rational requirement to be omniscient. So even though the omniscience plan is the best plan, there is no requirement to implement it after learning. As I'll put it, one is not required to be *plan coherent* with respect to it.

It should be clear that the problem with the omniscience plan is that it doesn't have the right *conditions*: by assigning different credence functions to every world, it makes distinctions that are too fine-grained for there to be a requirement to implement it after learning. To exclude such plans from consideration, we'll say that they are not *admissible* in the following sense:

Admissibility/Plan Coherence (definition): A plan is *admissible* if its condition is such that, after learning, a rational agent will implement the antecedently best actionable plan with that condition. We'll say that the agent is required to be *plan coherent* with

11. Formally, we can think of a practical act as a function from worlds to real numbers, which represent the utility of the act's outcome at that world.

12. That is, " $\phi \sim \psi$ " means that $\phi \succeq \psi$ and $\psi \succeq \phi$, and " $\phi \succ \psi$ " means that $\phi \succeq \psi$ and that it is not the case that $\psi \succeq \phi$.

13. Formally, the expected value of unconditional plan \mathcal{P} by the lights of credence function c is $EV_c(\mathcal{P}) =_{df} \sum_{\langle p, \phi \rangle \in \mathcal{P}} \sum_{w \in p} c(w) V(\phi, w)$, where $V(\phi, w)$ is the value of act ϕ at world w .

respect to the admissible plans.

Since we'll want to conduct our discussion both in terms of conditional and unconditional plans, I'll think of admissibility as a property of conditional plans in the first instance, and will say that an unconditional plan is admissible iff all of its elements are admissible.

Which plans are admissible? There are two natural answers to this question. On the first, one is not required to implement a plan when one has not learned its condition. In particular, if I have not learnt that w is actual, I am not required to implement the best plan for w , even though w is in fact actual. I am only required to implement the plan for the condition that I *have* learnt, i.e. the best plan for my evidence E . This gives us:

EPISTEMIC ADMISSIBILITY If the agent learns that E , the admissible plans are those with E as their condition, so that after learning that E , a rational agent will implement the antecedently best actionable plan for what to do or believe if E is true.¹⁴

The second answer is that we can exclude the omniscience plan because it is not a plan for what to believe given what one learns. That is, its conditions aren't propositions about what one has learnt ($\mathbb{L}E$ for some E), but cut across such propositions. In other words, there may be two worlds w_1 and w_2 such that the agent has learnt the same thing at those worlds, but where the omniscience plan recommends different beliefs. We can exclude such plans by assuming:

AUTO-EPISTEMIC (A-E) ADMISSIBILITY If the agent learns that E , the admissible plans are those with $\mathbb{L}E$ as their condition, so that after learning that E , a rational agent will implement the antecedently best actionable plan for what to do or believe if $\mathbb{L}E$

14. I remain neutral on whether the admissibility principles are narrow- or wide-scope norms. Whether **CONDITIONALIZATION**, as justified by the arguments in this paper, is narrow- or wide-scope depends on this choice. See Meacham (2015) for more discussion.

is true (i.e. if they learn that E).

The distinction between these two principles is easily overlooked, because they are equivalent in experiments with only transparent evidence. If any E that the agent may learn is transparent, then any such E is equivalent to $\mathbb{L}E$. Thus, the best plan for what to believe given E is just the best plan for what to believe given $\mathbb{L}E$.

What the standard arguments do is to focus on plans with E_1, E_2, \dots (or equivalently, $\mathbb{L}E_1, \mathbb{L}E_2, \dots$) as their condition, and prove that the plan to conditionalize is best among them. They can thus be understood as adopting one of these two accounts of admissibility, though this is not explicit in the presentation, because the attention is on identifying the best plan, not on what follows from the fact that it is best. That is, the standard arguments show:

PLAN CONDITIONALIZATION In any experiment with only transparent evidence, before learning, the plan to conditionalize on one's evidence is the best unconditional plan with $\mathbb{L}E_1, \mathbb{L}E_2, \dots$ (or equivalently, E_1, E_2, \dots) as its conditions. That is, $\text{PCond} \succ \mathcal{P}$ for any other doxastic unconditional plan \mathcal{P} with the same conditions.

This premise doesn't entail **CONDITIONALIZATION** on its own. Just because the plan to conditionalize on one's evidence was best *before learning*, it doesn't follow that later, after acquiring more evidence, one should implement it. We need something to bridge the gap between prior plans and posterior implementation. That is the job of the principles of admissibility we have just reviewed. Helping ourselves to one of these two principles (it doesn't matter which), we can now reason as follows: the best *conditional* plan for what to believe given E or $\mathbb{L}E$ must be to conditionalize on E , since, as **PLAN CONDITIONALIZATION** says, the best *unconditional plan* with $\mathbb{L}E_1, \mathbb{L}E_2, \dots$ (or equivalently, E_1, E_2, \dots) as its conditions says to conditionalize on E if they learn that E (or equiva-

lently, if E is true).¹⁵ Now applying either EPISTEMIC ADMISSIBILITY or A-E ADMISSIBILITY, it follows that upon learning that E , a rational agent will conditionalize on E , which is just what CONDITIONALIZATION says.

4.3 Two Arguments for Plan Conditionalization

Having seen how to argue from PLAN CONDITIONALIZATION to CONDITIONALIZATION, we will now look at two arguments for the former principle.

4.3.1 The Accuracy Argument

First, we have the *Accuracy Argument* by Hilary Greaves & David Wallace (2006),¹⁶ which forms part of the *Accuracy-First Epistemology* research program.¹⁷ The fundamental ideas of this program are, first, that the sole epistemic good is the *accuracy* of belief states — intuitively, their

closeness to the truth — and second that rational requirements on belief are justified insofar as conforming to them is conducive to this good. Whether a requirement is conducive to accuracy can be determined using decision-theoretic principles, in our case by comparing the expected accuracy of different plans.

How to measure accuracy is a difficult question, but Greaves & Wallace only make the standard assumption of *Strict Propriety*, which says that probability functions will regard themselves as expectedly more accurate than any other credence function.¹⁸ Using this assumption, Greaves & Wallace prove:

CONDI-MAX In experiments with only transparent evidence, the plan that uniquely maximizes expected accuracy among plans with $\mathbb{L}E_1, \mathbb{L}E_2, \dots$ as their conditions is the plan to conditionalize on one's evidence, i.e. PCond.

Combining this with the claim that rational preferences between unconditional plans are determined by their expected accuracy, PLAN CONDITIONALIZATION follows.

4.3.2 The Diachronic Dutch Book Argument

Our second argument is the *Diachronic Dutch Book Argument (DDBA)* for PLAN CONDITIONALIZATION (Lewis, 1999; Teller, 1973). A dutch book argument aims to establish an epistemic norm by showing that someone who doesn't conform to it is disposed to accept every one of a collection of bets that jointly risk the agent money without any corresponding chance of gain (a "dutch book"). This is thought to reveal an agent who violates such a norm as having incoherent, and therefore irrational, commitments. In the planning framework, this assumption amounts to:

NON-EXPLOITABILITY The best plan with $\mathbb{L}E_1, \mathbb{L}E_2, \dots$ as its conditions is one which doesn't lead the agent to accept a dutch

15. While this inference is intuitive, it can be made more rigorous by appeal to the principles presented in §§7.2, as follows: Suppose for reductio that, for some E_i , the plan to conditionalize on E_i if they learn E_i is not the best admissible plan, so that for some credence function $c^* \neq C(\cdot|E_i)$, it is not the case that $\langle \mathbb{L}E_i, C(\cdot|E_i) \rangle \succ \langle \mathbb{L}E_i, c^* \rangle$. Then, by CONDITIONAL COMPLETENESS, $\langle \mathbb{L}E_i, c^* \rangle \succeq \langle \mathbb{L}E_i, C(\cdot|E_i) \rangle$. Moreover, for every $i \neq j$, $\langle \mathbb{L}E_j, C(\cdot|E_j) \rangle \succeq \langle \mathbb{L}E_j, C(\cdot|E_i) \rangle$, by REFLEXIVITY. Hence, since $\mathbb{L}E_1, \dots, \mathbb{L}E_n$ is a partition (only one proposition can be the strongest claim that the agent has learnt), by CONDITIONAL DOMINANCE, we would have $\{\dots, \langle \mathbb{L}E_{i-1}, C(\cdot|E_{i-1}) \rangle, \langle E_i, c^* \rangle, \langle \mathbb{L}E_{i+1}, C(\cdot|E_{i+1}) \rangle, \dots\} \succeq$ PCond, so that PCond wouldn't be the best admissible unconditional plan, contrary to PLAN CONDITIONALIZATION. Hence our assumption must be false, and $\langle \mathbb{L}E_i, C(\cdot|E_i) \rangle$ must be best.

16. While Greaves & Wallace themselves aren't explicit about the fact that their argument is a plan-based one, this is made explicit in (Easwaran, 2013) and (Pettigrew, 2016, ch. 14).

17. See Pettigrew (2016) for an introduction and survey. I should note there are other accuracy-based arguments for CONDITIONALIZATION. One, namely (Leitgeb & Pettigrew, 2010) is not, I think, vulnerable to the criticisms that I describe here, although I think it has other problems (Pettigrew, 2016, ch. 15). Others — Easwaran (2013) and Briggs & Pettigrew (2020) — make similar assumptions to Greaves & Wallace's argument, and so I believe that what I say here carries over to their argument. In particular, both Easwaran and Briggs & Pettigrew assume that the agent's possible evidential propositions are guaranteed to be true and form a partition, assumptions which are equivalent to the claim that the agent can learn only opaque evidence (Schoenfield, 2017).

18. Formally, for any two distinct credence functions c, c^* , if c is a probability function, then $Ea_c c > Ea_c c^*$, where a is the accuracy measure. Greaves & Wallace (2006, p. 620) use the term "everywhere strongly stable" for a strictly proper accuracy measure.

book.

As it turns out, there is a unique such plan, namely the plan to conditionalize on one's evidence, as the following result establishes:

(CONVERSE) DUTCH BOOK THEOREM FOR CONDITIONALIZATION In experiments with only transparent evidence, if a rational agent follows a doxastic plan with $\mathbb{L}E_1, \mathbb{L}E_2, \dots$ as its conditions, they are disposed to accept a dutch book over time if and only if they don't follow the plan to conditionalize.

From these two premises, PLAN CONDITIONALIZATION follows: the best admissible plan is one which avoids dutch-bookability, and there is exactly one such plan, namely the plan to conditionalize on one's evidence.¹⁹

5. A Dilemma for the Externalist

As noted, the standard arguments for CONDITIONALIZATION are restricted to experiments where all potential bodies of evidence are transparent. If we endorse EVIDENTIAL TRANSPARENCY, no other experiments are possible, so the arguments establish that one should always conditionalize. However, on this assumption, it would follow that EXTERNALISM is false. If one conditionalizes on a proposition, one becomes certain in it and in all equivalent propositions. So as long as our agent conditionalizes on their evidence, as our two arguments say they should if we assume EVIDENTIAL TRANSPARENCY, they will become certain of what it is that they have learnt, since transparent evidence by definition is equivalent to the claim that it is their evidence.

What if, on the other hand, we deny EVIDENTIAL TRANSPARENCY? Then CONDITIONALIZATION is *consistent* with EXTERNALISM: condition-

alizing on opaque evidence can leave one uncertain of what one's evidence is. For example, if in *Here's a Hand* you are in the Bad Case, you learn *EXP*, which is true in both the Good and the Bad Cases. Thus, conditionalizing on one's evidence will leave you uncertain of whether you are in the Good or the Bad Case. And since your evidence is different in those two cases, you will also be uncertain of what your evidence is.

An argument for CONDITIONALIZATION that covers experiments with opaque evidence would therefore be just what the externalist needs. However, generalizing the plan-based arguments to such experiments raises the question of what our account of *admissibility* should be: EPISTEMIC ADMISSIBILITY or A-E ADMISSIBILITY. While, as we saw, these principles are equivalent in experiments with only transparent evidence, this is no longer the case when opaque evidence is possible. If one is in the Bad Case, one's evidence *EXP* is not equivalent to the claim that one has learnt *EXP*. Thus, the best plan for what to believe given *EXP* need not be the same as the best plan for what to believe given $\mathbb{L}EXP$.

What several recent papers have done is to generalize the two arguments by focusing on plans with the $\mathbb{L}E$ s as their condition, which in our terms amounts to adopting A-E ADMISSIBILITY. To be clear, because the distinction between the two principles of admissibility is overlooked, this choice is not explicitly made. Rather, this account of admissibility is simply assumed as the natural way to block plans such as the omniscience plan from being considered. Given this assumption, Matthias Hild (1998) has shown that the (CONVERSE) DUTCH BOOK THEOREM FOR CONDITIONALIZATION doesn't generalize to experiments with potentially opaque evidence. Similarly, Miriam Schoenfield (2017) shows that CONDI-MAX doesn't generalize to such experiments either.²⁰ That is not all, however: Hild and Schoenfield also show that in experiments with

19. While not all treatments of the *DDBA* interpret it as a plan-based argument, most philosophers who explicitly consider the distinction between CONDITIONALIZATION and PLAN CONDITIONALIZATION agree that it can only establish the former, if at all, via the latter. For particularly clear statements of this claim, see Christensen (1996) and Pettigrew (2016, p. 188).

20. Similar conclusions are reached by Bronfman (2014) in the case of the *Accuracy Argument* and by Gallow (2019) in the case of the *DDBA*. Das (2019) proves a result similar to Schoenfield's in a framework where the agent updates their *ur-prior* on their *total* evidence. Though I will not attempt to show it here, I believe that my argument applies to Das's framework, too.

possibly opaque evidence, the best plan is the plan to follow A-E CONDITIONALIZATION. That is, the plan that maximizes expected accuracy, and the only plan that doesn't render the agent exploitable by dutch books, is the plan to comply with the following update rule:

AUTO-EPISTEMIC (A-E) CONDITIONALIZATION Upon learning that E , a rational agent will conditionalize their prior credences on $\mathbb{L}E$, so that, for any proposition p ,

$$C_E(p) = C(p|\mathbb{L}E) =_{\text{df}} \frac{C(p \wedge \mathbb{L}E)}{C(\mathbb{L}E)},$$

if defined, where C and C_E are the agent's credence functions before and after learning that E , respectively.

Recall that CONDITIONALIZATION doesn't just say to conditionalize on *some proposition*, but specifically on one's evidence. A-E CONDITIONALIZATION agrees with the first part of that claim, but not with the second. It says to conditionalize, not on one's evidence (call it E), but on the true description of one's evidence ($\mathbb{L}E$). Note that if E is transparent, the agent is certain in advance that E will be their evidence iff E is true. Hence, conditionalizing on E will have the same effect as conditionalizing on $\mathbb{L}E$, since conditionalizing on equivalent propositions yields equivalent results. Thus, under the restriction to transparent evidence, A-E CONDITIONALIZATION and CONDITIONALIZATION always agree. But if we reject EVIDENTIAL TRANSPARENCY and thus accept that opaque evidence is possible, the two rules can come apart.

In other words, what Hild and Schoenfield have shown is that if the restriction to experiments with only transparent evidence is lifted, it is the plan to *auto-epistemically* conditionalize that has the nice properties: it maximizes expected accuracy among plans with $\mathbb{L}E_1, \mathbb{L}E_2, \dots$ as their conditions and is the only such plan that renders the agent immune to exploitability by dutch books. Combining these facts with the rest of

the premises of the *Accuracy* and *Dutch Book Arguments*, we get:

PLAN AUTO-EPISTEMIC (A-E) CONDITIONALIZATION Before learning, the plan to auto-epistemically conditionalize is the best plan with $\mathbb{L}E_1, \mathbb{L}E_2, \dots$ as its conditions. That is,

$$\{\langle \mathbb{L}E_1, C(\cdot|\mathbb{L}E_1) \rangle, \langle \mathbb{L}E_2, C(\cdot|\mathbb{L}E_2) \rangle, \dots\} \succ \mathcal{P}$$

for any other doxastic unconditional plan \mathcal{P} with the same conditions.

This premise can then be combined with A-E ADMISSIBILITY, just as before, to get an argument for A-E CONDITIONALIZATION itself.

Now, here's the bad news for the externalist: A-E CONDITIONALIZATION is inconsistent with EXTERNALISM. One becomes certain of what one conditionalizes on, so if one conditionalizes on the true description of one's evidence, naturally one will never be uncertain of what one's evidence is! For example, in *Here's a Hand*, if one is in the Bad Case, A-E CONDITIONALIZATION will say to conditionalize on $\mathbb{L}EXP$ — the claim that the strongest proposition that one learnt is that one had an experience as of a hand. This will leave one certain that one is in the Bad Case. So rejecting EVIDENTIAL TRANSPARENCY will not help the externalist: we still cannot endorse any of the two arguments as sound and retain EXTERNALISM.

To sum up: the result is a dilemma for the externalist. On the first horn, they can assume EVIDENTIAL TRANSPARENCY. Then the two arguments will give them CONDITIONALIZATION, but not EXTERNALISM. On the second horn, they can reject EVIDENTIAL TRANSPARENCY. That's what the externalist wanted to do all along — if it's possible to be uncertain about evidential matters, that's surely because one can have opaque evidence. But then the premises of the two arguments will support A-E CONDITIONALIZATION, which again is inconsistent with EXTERNALISM. So regardless of whether they accept EVIDENTIAL TRANSPARENCY, the premises of our two arguments will lead them to reject EXTERNALISM.

6. Rethinking Plan Coherence

6.1 Which Plans Are Admissible?

How can the externalist respond? They could reject the accuracy framework, and the idea that dutch-bookability indicates irrationality. My suggestion, however, is to reconsider something that they have in common: the assumption about plan coherence (A-E ADMISSIBILITY) that has to be made in order to go from PLAN A-E CONDITIONALIZATION to A-E CONDITIONALIZATION itself. In particular, I think that externalists have good reason to reject A-E ADMISSIBILITY in favour of EPISTEMIC ADMISSIBILITY. That is, the conditions of the admissible plans are what you've learnt, not that you've learnt it. Since A-E ADMISSIBILITY is a premise in the arguments for A-E CONDITIONALIZATION, this allows externalists to reject those arguments as unsound.

To see why, let us think about practical planning for a moment. Imagine that our agent is now playing poker. Before the next card is drawn, they rightly conclude that, if they have a better hand than their opponents, the plan to raise the bet is best. The card is revealed, and as a matter of fact they do now have the best hand. However, since they cannot see their opponents' hands, they don't know this. In fact, given what the agent knows about the cards on the table and their opponents' behaviour, it now looks like they have a worse hand than some of their opponents. In this situation, they are obviously not required to raise the bet, despite the fact that the best plan says to do so. The natural explanation, I suggest, is that they have not learnt that they have the best hand, and so are not required to implement the best plan for that condition. This is precisely the explanation offered by EPISTEMIC ADMISSIBILITY, according to which a requirement to implement the best plan in a class only takes effect when one has learned the condition of those plans.

Now returning to the epistemic case, in *Here's a Hand*, this explanation shows where A-E ADMISSIBILITY goes wrong. In the Bad Case, while you have learnt *EXP*, you have not learnt *LEXP*. Thus, the same objection as in the poker case applies: you have not learnt the condition

of the best *LEXP*-plan, and so aren't required to implement it. To put the point differently: everyone will acknowledge that if we count one's "situation" as what hand one has in a poker game, one should not always prefer what the better plan for one's situation says. My suggestion is that this is because everyone agrees that one can have the best hand in a poker game, without having learnt this. But if we think of one's "situation" as a description of what one's evidence is, it has proved more tempting to think that one *is* required to prefer what the better plan for one's situation says. However, if opaque evidence is possible, the same objection applies here: one can learn that *E*, without having learnt that one learnt that *E*.²¹

This argument against A-E ADMISSIBILITY from EPISTEMIC ADMISSIBILITY can be put on theoretically firmer footing by showing that the latter follows from two plausible principles of practical reasoning. First we have a diachronic principle:

IRRELEVANCE If what a rational agent learns doesn't rule out any *p*-worlds, their preferences between plans with *p* as their condition will remain unchanged.

The idea is that if the agent doesn't learn anything that rules out *p*-worlds, the new evidence doesn't put them in a position to reconsider their plans for *p*, since they have not learnt anything about what could happen if *p* is true. The second premise concerns the agent's practical reasoning after learning:

IMPLEMENTATION If, after learning, a rational agent prefers $\langle p, \phi \rangle$ to all other actionable plans with *p* as their condition, and if their evidence entails that *p*, then they will ϕ .

This is a synchronic principle of practical rationality, akin to that of means-end reasoning. We can think of it as saying that one can reason

²¹. This will be the case both in failures of Positive and Negative Access (see Note 7), since "learns" and "evidence" stand for the *strongest* proposition that the agent has learnt.

from one's evidence and one's preferences between conditional plans to a decision to implement the preferred plan. For example, if my best plan for rain is to bring a rain coat rather than an umbrella, and my evidence entails that it is raining, then I can decide to bring the rain coat on the basis of those two attitudes.

EPISTEMIC ADMISSIBILITY follows from these two premises. First, by learning that E , a rational agent doesn't learn anything that rules out any E -worlds. Hence, their preferences between E -plans will remain unchanged, so that they still prefer the antecedently best E -plan to other E -plans. Second, if the agent learns that E , they have learnt something that entails that E , so that they will implement the best actionable E -plan, which as just shown is the *antecedently* best actionable E -plan.

Importantly, there is no parallel argument for A-E ADMISSIBILITY. Consider the best $\mathbb{L}E$ -plan. To establish A-E ADMISSIBILITY, we would want to reason in two steps, as before: first that the agent's preferences between $\mathbb{L}E$ -plans will stay the same upon learning that E , and second that the agent will implement the best $\mathbb{L}E$ -plan after learning. For the sake of argument, I am happy to grant the first step, since it follows from IRRELEVANCE on the assumption that the agent will only learn E if it is true.²² On this assumption, learning that E cannot rule out any $\mathbb{L}E$ -worlds, since they are also E -worlds, so that IRRELEVANCE says that the agent's preferences between $\mathbb{L}E$ -plans will remain the same upon learning that E .

However, we would then have to show that the agent should implement the best $\mathbb{L}E$ -plan after learning. In cases of opaque evidence, this doesn't follow from IMPLEMENTATION: in the Bad Case, the agent doesn't learn anything that entails $\mathbb{L}EXP$ by learning that EXP , and so it doesn't follow that they are required to implement the best plan

for this condition. In other words, if the agent's evidence is opaque, they cannot necessarily reason "I have learnt that E , and the best plan for learning that E is to ϕ , so I'll ϕ ", because they have not learnt that they have learnt that E , and so cannot use $\mathbb{L}E$ in their deliberation. This is just an instance of the general fact that rational requirements are triggered not by just any facts, but by *learnt* facts. For example, if I know that *if p , then q* , this doesn't commit me to believing that q unless I also know (or *learn*) that p . Similarly, the fact that I prefer the plan to ϕ to the plan to ψ *given some condition* doesn't trigger a rational requirement to prefer ϕ to ψ unless I have learnt the condition.

Finally, recall that we introduced the admissibility principles to exclude plans such as the omniscience plan. Arguably, however, the stated motivations for excluding this and similar plans speak against A-E ADMISSIBILITY in cases of opaque evidence. First, Greaves & Wallace say that plans such as the omniscience plan require the agent to "respond to information that he does not have" (Greaves & Wallace, 2006, p. 612). I think that this problem is sometimes also shared by $\mathbb{L}E$ -plans: requiring the agent to auto-epistemically conditionalize in the Bad Case is precisely to require them to "respond to information which he does not have" — namely the information that they are in the Bad Case. Even more revealing is Kenny Easwaran's (2013, pp. 124–125) motivation for excluding the omniscience plan: that a plan should assign the same credence function to two states if the agent "does not learn anything to distinguish them". Again, this problem is shared by $\mathbb{L}E$ -plans when evidence is opaque: in the Bad Case the agent doesn't learn anything to distinguish the Good Case from the Bad Case, and so should not be required to act on the plan to conditionalize on being in the Bad Case if they are in the Bad Case.

6.2 Is My Argument Unexternalist?

One may worry that, while the case against A-E ADMISSIBILITY seems reasonable, it clashes with a fundamental externalist principle: that facts about rationality can fail to be known to the agent. Doesn't my

22. The assumption that evidence is factive is widespread in formal epistemology, including among externalists (Littlejohn, 2011; Williamson, 2000). It is not without critics, however (Arnold, 2013; Comesana & McGrath, 2016; Comesaña & Kantin, 2010; Rescorla, 2021; Rizzieri, 2011). Such critics will have an additional reason to reject A-E ADMISSIBILITY: perhaps learning that E can warrant a reconsideration of one's plan for $\mathbb{L}E$.

complaint that one is not required to implement a plan if one doesn't know its condition contradict this principle? This would be a serious problem if true, since my aim is to render Bayesianism consistent with EXTERNALISM.

In reply, it is important to distinguish between *rules* and *plans*. The former are principles of rationality, and the objection is right that an externalist wouldn't want to say that such a principle only applies when the agent knows what it says to do. My objection, however, is about the *plan*. A plan is a commitment of the agent, like a belief or an intention.²³ What I have claimed is that this commitment only generates the additional commitment to implementing the plan when the agent has learnt its condition.

This is perfectly in line with what externalists think. No externalist would say that any fact can be rationally used in one's reasoning just by virtue of being true. To return to the above example, no externalist would hold the absurd view that if one knows *if p then q*, and *p* is true but not known or believed to be, one can rationally use *p* in one's reasoning and is thereby rationally committed to believing that *q*. Rather, they think that one must stand in some substantive epistemic relation to *p*, such as *knowing it* or *being in a position to know it*, in order for such a requirement to be triggered. Or to take a practical example, my intention or desire to drink gin doesn't rationally require me to intend or desire to drink the contents of the glass just because it happens to contain gin. A requirement is only triggered if I *know* that the glass contains gin, even for an externalist. That's why many externalists think of one's evidence as well as one's practical reasons as consisting of the propositions that one *knows*, or stands in similar relations to, such as what one *non-inferentially knows* or is in a *position to know* (Littlejohn, 2011; Lord, 2018; Williamson, 2000). In the Bayesian framework, we call this relation "learning" or "having the proposition as evidence".

The effects of the externalist principle that requirements of rationality

23. To be more precise, the agent commits themselves by *preferring* the plan to other plans with the same condition.

can fail to be known are felt on another level. One source of uncertainty about rationality is that that one can stand in the right epistemic relation to a proposition without standing in the right relation to the higher-order proposition that says that one stands in the right relation to that proposition. So, for example, one can know *if p then q* and also know *p*, but fail to know that one knows that *p*. In such a case, the externalist would say that one is required to believe that *q*, even though one is not in a position to know that this is required. Similarly, an externalist about evidence would say that one can sometimes learn some *E* without learning $\perp E$. It would then follow from EPISTEMIC ADMISSIBILITY that one is required to implement the best plan for *E* even though one doesn't know that one is so required.

7. Foundations for New Arguments

7.1 Why Plan Coherence?

I have argued that we should reject A-E ADMISSIBILITY in favour of EPISTEMIC ADMISSIBILITY. This suggests the following strategy: to argue for an update rule, we should consider plans with the evidence rather than a description of the evidence as their condition. In particular, if we could show that the plan to adopt $C(\cdot|E)$ is the best *E*-plan, EPISTEMIC ADMISSIBILITY would then kick in and require the agent to adopt the conditionalized credences. This is the line that I will pursue in the rest of this paper.

Before I do so, I want to address a more general challenge: I have argued that EPISTEMIC ADMISSIBILITY is a more plausible expression of the idea that rationality requires one to "stick to the plan" than A-E ADMISSIBILITY. But why think that rationality requires plan coherence of any kind? That is, why think that there are any admissible plans at all? I have already argued that if one did not learn the condition of the best plan, one need not be required to implement it. One can also have learnt *too much* to be required to implement it, as in a case where I have planned to sell my lottery ticket if I am offered \$100 for it, but then learn that I am offered this amount *and* that it is the winning ticket.

Here I have learnt *more* than the condition of the plan, which puts me in a position to rationally re-evaluate it.

What *is* plausible is that one should stick to a plan if one has learnt *just enough*, i.e. if one has learnt *exactly* the condition of the plan, no more and no less. Admittedly, I don't have an argument against general scepticism about plan coherence,²⁴ though in this respect I am not in a worse position than those who appeal to A-E ADMISSIBILITY. To the sceptic, all I can offer is the intuitive plausibility of the claim that one should stick to the plan when none of the above objections apply. This is best brought out by considering how someone would justify their violation of the principle:

A: Why did you bring your umbrella? I thought you said yesterday that wearing a rain coat was the better plan for rain.

B: It was, yes. But then I learnt something that made me reconsider. Yesterday night I didn't know that it was going to rain, so bringing the rain coat was the better plan for rain. But this morning, I learnt that it was raining, and in the light of that new information I decided to bring the umbrella instead.

A: Wait, how can the fact that it is raining make you revise your plan about what to do *if it rains*?

A's reaction seems justified. Either B was wrong about the plan to begin with, or they were wrong to revise it. That something happened cannot itself be a reason to reconsider what to do if it happens.

7.2 Conditional Plan Evaluation

In order to develop an argument for CONDITIONALIZATION based on EPISTEMIC ADMISSIBILITY, an adjustment will be necessary. The original arguments, which are restricted to experiments with only transparent evidence, establish CONDITIONALIZATION by showing that the best conditional plan for learning that E (i.e. for $\mathbb{L}E$) is to conditionalize on E .

24. In terms of the argument in §§6.1, this kind of sceptic would disagree with IRRELEVANCE rather than IMPLEMENTATION.

This claim was in turn justified by arguing that the best *unconditional* plan with $\mathbb{L}E_1, \mathbb{L}E_2, \dots$ as its conditions was the plan to conditionalize on E_1 if the agent learns that E_1 , conditionalize on E_2 if they learn that E_2 , etc.

Since we intend to use plans with E_1, E_2, \dots as their conditions in a setting with opaque evidence, this kind of justification is not available, however. One can only form an unconditional plan from conditional plans if their conditions form a partition, but E_1, E_2, \dots is not guaranteed to form a partition in experiments with opaque evidence. For example, in *Here's a Hand*, one proposition that you might learn is EXP , and another is $EXP \wedge HAND$. These propositions are consistent, and thus a plan to believe one thing given EXP and another given $EXP \wedge HAND$ is not a real unconditional plan. However, as I will show, there is a similar way of evaluating conditional plans that escapes this problem. To this end, I will assume three principles about rational preference between conditional plans, and show that, combined with the claim that one ought to maximize expected utility, they pin down a unique way of evaluating conditional plans.

First, it is highly plausible that the preference relation between conditional plans satisfies some very minimal structural properties — that it is reflexive, and that two conditional plans with the same condition can always be compared:

REFLEXIVITY For any conditional plan A , $A \succeq A$.

CONDITIONAL COMPLETENESS For any condition p and acts ϕ, ψ , either $\langle p, \phi \rangle \succeq \langle p, \psi \rangle$ or $\langle p, \psi \rangle \succeq \langle p, \phi \rangle$ (or both).²⁵

The third principle will do the heavy lifting, and is a version of Leonard Savage's (1954) *Sure-Thing Principle*.²⁶ The idea is that if you have some

25. This principle is not needed to prove the CONDITIONAL EXPECTATION LEMMA below, but will be used in the next section of the paper.

26. See Maher (1993, p. 10) for a discussion of the different senses of the term "The Sure-Thing Principle".

partition of logical space, and for each cell of the partition prefer some act to another, you should be able to combine those preferences between conditional plans to form a preference between the *unconditional* plans that they make up. For example, suppose that it is better to bring an umbrella than a rain coat if it rains, and better to bring sunglasses than a cap if it doesn't rain. If so, the unconditional plan to bring an umbrella if it rains and sunglasses if it doesn't must be better than the plan to bring a rain coat if it rains and a cap if it doesn't. Formally:

CONDITIONAL DOMINANCE For any acts $\phi_1, \dots, \phi_n, \psi_1, \dots, \psi_n$, and any partition p_1, \dots, p_n , if for all i ($1 \leq i \leq n$),

$$\langle p_i, \phi_i \rangle \succeq \langle p_i, \psi_i \rangle$$

then

$$\{\langle p_i, \phi_i \rangle | 1 \leq i \leq n\} \succeq \{\langle p_i, \psi_i \rangle | 1 \leq i \leq n\}.$$

These assumptions, combined with the idea that we evaluate unconditional plans by their expected value, suffice to establish the following claim:

CONDITIONAL EXPECTATION LEMMA Given CONDITIONAL DOMINANCE and REFLEXIVITY, and if preferences between unconditional plans are determined by their expected value, then for any acts ϕ, ψ (whether doxastic or practical), and any proposition E , if $C(\cdot|E)$ is defined and

$$\langle E, \phi \rangle \succeq \langle E, \psi \rangle$$

then

$$EV_{C(\cdot|E)}\phi \geq EV_{C(\cdot|E)}\psi$$

where $EV_c A$ is the expected value of act A according to credence function c .

The proof is in the [Appendix](#). We have thus arrived at a way of evaluating conditional plans: rational preferences between plans with some condition reflect their conditional expected value.

8. Two Arguments for Conditionalization

We can now combine this close connection between conditional credences and preferences about conditional plans with EPISTEMIC ADMISSIBILITY to give two new arguments for CONDITIONALIZATION: one accuracy-theoretic, and one practical. These arguments are not restricted to experiments with transparent evidence. As such, their conclusion is consistent with EXTERNALISM: if it is possible to have opaque evidence, they require the agent to conditionalize on it, and thus to become uncertain of what they have learnt.

8.1 A New Accuracy Argument

I begin with the accuracy-theoretic argument, which builds on Greaves & Wallace's *Accuracy Argument*, but which is now stated in terms of the best conditional doxastic plan for the agent's evidence. Just as before, it proceeds in two steps. First, it is shown that the *conditional* plan to conditionalize on E is the best plan for E :

CONDITIONAL PLAN CONDITIONALIZATION For any possible evidence E , the best plan for E is to conditionalize on E . That is, for any credence function $c^* \neq C(\cdot|E)$, $\langle E, C(\cdot|E) \rangle \succ \langle E, c^* \rangle$.²⁷

Justification: Assume, for reductio, that there is some credence function $c^* \neq C(\cdot|E)$ such that it is not the case that $\langle E, C(\cdot|E) \rangle \succ \langle E, c^* \rangle$. By CONDITIONAL COMPLETENESS, it follows

27. Since the simplifying assumptions in §2 guarantee that $C(\cdot|E)$ is defined for any E , I omit the "...if $C(\cdot|E)$ is defined" qualifier in my presentation of the two arguments.

that $\langle E, c^* \rangle \succeq \langle E, C(\cdot|E) \rangle$. Now applying the CONDITIONAL EXPECTATION LEMMA, we get $EV_{C(\cdot|E)}c^* \geq EV_{C(\cdot|E)}C(\cdot|E)$. Assuming that the value of a credence function is its accuracy, it follows that $C(\cdot|E)$ regards c^* as at least as accurate as itself in expectation. But that is impossible, since by *Strict Propriety* any probability function regards itself as more accurate in expectation than any other credence function, and $C(\cdot|E)$ is a probability function, since C is. Hence the assumption must have been false: $\langle E, C(\cdot|E) \rangle$ is the best doxastic conditional plan with E as a condition.

Note that this principle is about the best plan for E , not the best plan for $\mathbb{L}E$. We can thus argue as follows: Suppose that our rational agent learns that E . Assuming that all doxastic plans are actionable (i.e. that the agent can adopt any credence function upon learning that E), it follows from CONDITIONAL PLAN CONDITIONALIZATION that the best actionable E -plan is to conditionalize on E . Thus, by EPISTEMIC ADMISSIBILITY the agent will conditionalize on E . Hence, CONDITIONALIZATION is true. Since no assumptions have been made about the relationship between E and $\mathbb{L}E$, this conclusion holds irrespective of whether E is transparent or opaque.

8.2 *A Practical Argument*

The CONDITIONAL EXPECTATION LEMMA also suggests a different argument for CONDITIONALIZATION, one that doesn't employ the accuracy framework. In fact, we can dispense with the idea of doxastic planning altogether. As long as *practical planning* — planning what to do, rather than what to believe — obeys the assumptions listed in the CONDITIONAL EXPECTATION LEMMA, we can establish CONDITIONALIZATION by

combining EPISTEMIC ADMISSIBILITY with expected utility reasoning.²⁸

Acts are now practical acts, instead of doxastic acts, and we'll assume that preferences between them are determined by their expected utility. The *Practical Argument* for CONDITIONALIZATION then goes as follows:

Suppose, for reductio, that CONDITIONALIZATION is false, so that for some E and A , $C_E(A) \neq C(A|E)$. Assume that our agent can perform one of the following two acts, where acts are functions from worlds to the utility, in that world, of performing the act:

$$\mathcal{A}(w) = \begin{cases} 1 - C_E(A) & \text{if } w \in A \\ -C_E(A) & \text{if } w \notin A \end{cases}$$

and

$$\mathcal{O}(w) = 0 \text{ for all } w.$$

We can think of \mathcal{A} as accepting a \$1 bet on A for the price of $C_E(A)$, and of \mathcal{O} as declining to bet.

Since $C_E(A) \neq C(A|E)$, it follows that either $EV_{C(\cdot|E)}\mathcal{A} > EV_{C(\cdot|E)}\mathcal{O}$ or $EV_{C(\cdot|E)}\mathcal{A} < EV_{C(\cdot|E)}\mathcal{O}$.²⁹ Thus, by the CONDITIONAL EXPECTATION LEMMA, it is not both the case that (a) $\langle E, \mathcal{A} \rangle \succeq \langle E, \mathcal{O} \rangle$ and (b) $\langle E, \mathcal{O} \rangle \succeq \langle E, \mathcal{A} \rangle$. However, by CONDI-

28. Another alternative would be to take a hybrid approach: think about doxastic rather than practical planning, but evaluate doxastic plans by the practical utility of the decisions they justify. This is the approach of Peter M. Brown's (1976) argument for CONDITIONALIZATION. Just as the other arguments I have discussed, Brown's argument is restricted to experiments with transparent evidence, but can be adapted, in the way I have adapted Greaves & Wallace's argument, to rid it of this restriction and render it consistent with EXTERNALISM, though showing this is beyond the scope of this paper.

29. $EV_{C(\cdot|E)}\mathcal{A} = C(A|E)(1 - C_E(A)) + (1 - C(A|E))(-C_E(A)) = C(A|E) - C_E(A)$, which is either greater or smaller than $EV_{C(\cdot|E)}\mathcal{O} = 0$ since $C_E(A) \neq C(A|E)$.

TIONAL COMPLETENESS, one of (a) or (b) is true, so that by definition of strict preference (\succ) it follows that either $\langle E, \mathcal{A} \rangle \succ \langle E, \mathcal{O} \rangle$ or $\langle E, \mathcal{O} \rangle \succ \langle E, \mathcal{A} \rangle$.

Now applying EPISTEMIC ADMISSIBILITY, it follows that the agent must either perform \mathcal{A} or \mathcal{O} after learning that E . But \mathcal{A} is a fair bet from the point of view of C_E (i.e. \mathcal{A} and \mathcal{O} have the same expected utility for the agent after learning that E)³⁰ so that the agent is permitted to perform either. Thus, our assumption must be wrong: it cannot be the case that for some E, \mathcal{A} , $C_E(\mathcal{A}) \neq C(\mathcal{A}|E)$. CONDITIONALIZATION must be true.

Two remarks on this argument: First, just like the *New Accuracy Argument*, it is not restricted to experiments with transparent evidence. Second, note that by applying EPISTEMIC ADMISSIBILITY, we implicitly assumed that the agent has exactly two available acts: \mathcal{A} and \mathcal{O} . How does this justify CONDITIONALIZATION in a situation where those acts aren't available, or where others are? Here we can take a page from the *DDBA*, which is meant to show that one should always conditionalize on one's evidence, not just when a clever bookie is in the vicinity and the only available options are to accept their bet, or not. Proponents of the *DDBA*, such as David Lewis, argue that the fact that the agent *would* accept a dutch book if they faced that choice means that they have "contradictory opinions about the expected value of the very same transaction" (Lewis, 1999, p. 405), which is a kind of irrationality, not in the agent's actions, but in their *preferences*. The same can be said about this argument: the fact that the agent is indifferent between \mathcal{O} and \mathcal{A} , even though before learning they strictly preferred one of $\langle E, \mathcal{O} \rangle$ and $\langle E, \mathcal{A} \rangle$ to the other, shows that they have plan incoherent preferences even though they aren't able to perform those acts, or are able to

30. $EV_{C_E} \mathcal{A} = C_E(\mathcal{A})(1 - C_E(\mathcal{A})) + (1 - C_E(\mathcal{A}))(-C_E(\mathcal{A})) = 0 = EV_{C_E} \mathcal{O}$.

perform others.³¹

8.3 Discussion

My two arguments can be seen, roughly, as generalizations of the standard arguments to experiments that may contain opaque evidence, though the *Practical Argument* is more loosely related to the *DDBA*. They generalize the arguments in a different way than Hild and Schoenfield do, though, by adopting a different account of admissibility. Those who, like Hild and Schoenfield, consider the admissible plans to be those with $\mathbb{L}E$ as their condition, are really answering the question: What is the optimal way to react to the fact that one has learned that E ? My arguments, on the other hand, answer a different question: What is the optimal way to react to the fact that E is true?. It should be no surprise that our answers differ accordingly: I think one ought to conditionalize on E , while they conclude that one ought to conditionalize on $\mathbb{L}E$.

This doesn't, I think, reflect well on the arguments for $\mathbb{A}\text{-}E$ CONDITIONALIZATION. Arguably, we should want our update rule to help explicate the idea that one should believe what one's evidence supports. If so, what should matter is the evidential proposition itself (E), not the fact that it is one's evidence ($\mathbb{L}E$). That is because, intuitively, whether E supports p depends on the logical or inductive relationship between E and p , not on the logical or inductive relationship between $\mathbb{L}E$ and p . For example, *It is raining* doesn't intuitively support *I have evidence about the weather* to any high degree, given normal background evidence, because there is no logical or strong inductive connection between the fact about the weather and the fact about my mind. That is the case even though *My evidence says that it is raining* entails the claim that *I have evidence about the weather*.³²

31. A more careful presentation of this version of this argument would involve reformulating EPISTEMIC ADMISSIBILITY in terms of rational *preference*, instead of having the "actionable"/"able to perform" qualifier. The principle would then say that the agent is required to prefer ϕ to ψ if they learnt that E and antecedently preferred $\langle E, \phi \rangle$ to $\langle E, \psi \rangle$.

32. Barnett (2016) surveys this similar objections to transparency accounts of self-knowledge.

As for the differences between my arguments, each has its own advantage. The *Practical Argument*, I believe, has an important advantage over the *New Accuracy Argument*. The idea that we should think of the rationality of epistemic principles as determined by their conduciveness to accurate belief is highly controversial.³³ The *Practical Argument* allows us to sidestep this controversy, since the only assumption it makes about rational credences (apart from them being probabilities) is that they are connected to preferences about acts and plans via expected value calculations. This makes the *Practical Argument* interesting independently of the issue of EXTERNALISM. On the other hand, if one thinks that facts about epistemic rationality are explanatorily prior to facts about practical rationality, the *New Accuracy Argument* seems better poised to *explain*, rather than just establish, the rationality of conditionalizing.

8.4 Comparison to Another Solution

Dmitri Gallow (2021) suggests a different way of reconciling the *Accuracy Argument* with EXTERNALISM.³⁴ He gives an accuracy-based argument for an update rule which is consistent with EXTERNALISM, but differs from both CONDITIONALIZATION and A-E CONDITIONALIZATION.³⁵ His key idea is that one can evaluate plans not by the expected accuracy of correctly following them, as I have done, following most of the literature, but rather by the expected accuracy of *adopting* the plan, where the latter takes into account the possibility that one will not follow the plan correctly.³⁶ He shows that if expected accuracy is calculated in this way, his update rule maximizes expected accuracy.

33. See Caie (2013), Carr (2017), Greaves (2013), and Konek & Levinstein (2019).

34. Thanks to a reviewer for suggesting that I compare my view to Gallow's.

35. Gallow's terminology differs from mine: my EXTERNALISM he calls "Certainty Externalism", while the principle he calls "Externalism" is the view that one can have evidence that doesn't tell one what one's evidence is. An externalist in his sense is thus a denier of EVIDENTIAL TRANSPARENCY, in my terminology.

36. Strictly speaking, Gallow argues for the claim that one should be *disposed*, rather than *plan*, to conditionalize. My presentation here transposes his argument into to the key of planning.

While our update rules differ, this doesn't necessarily mean that we disagree across the board. Instead, they are different kinds of rules: Gallow's rule concerns a less idealized standard of rationality, which takes possible fallibility into account, while my rule is about *ideal* rationality, which is only sensitive to the agent's lack of information, and not to the possibility that they are cognitively limited.³⁷ We do disagree about the ideal case, however. As Gallow acknowledges, his rule collapses into A-E CONDITIONALIZATION for ideal agents who know that they will correctly follow their plan.³⁸ For Gallow, then, such agents can never be uncertain of their evidence, even when it is opaque as in *Here's a Hand*. Thus, I reject Gallow's argument for the same reason that I reject Hild's and Schoenfield's arguments: in cases of opaque evidence the agent has not learnt the condition of their update plan and so is not required to implement it. My view thus has an advantage as far as the externalist is concerned. For me one can be rationally uncertain of one's evidence for the same reason one can be rationally uncertain about anything else: one's evidence may not settle the matter. For Gallow this is not enough for uncertainty about evidence to be rational. One must also be uncertain about one's capacity to implement plans.

8.5 Is Conditionalization Still a Coherence Norm?

I end by considering a more general worry about the relationship between EXTERNALISM and Bayesianism.³⁹ Standard Bayesian arguments such as accuracy or dutch book arguments are often thought to show that an agent who doesn't conform to the standard Bayesian norms is *incoherent* in that they are doing something that's suboptimal *from their own perspective*, for example, by accepting a Dutch Book, or by having beliefs that are accuracy-dominated. But if one can have opaque evidence, conditionalizing can't always be a matter of doing what's best from one's perspective, since in the Bad Case, for example, one doesn't

37. Here I am following Schoenfield (2015), who suggests that these two ways of evaluating plans correspond to two different meanings of "rational".

38. See Gallow (2021, p. 501).

39. Thanks to an anonymous reviewer for suggesting this worry.

know whether it is best to conditionalize on EXP or $EXP \wedge HAND$.

In reply, an externalist will insist on the distinction between what's best from one's perspective and what one can *know* (or be certain) is best from one's perspective. My arguments do show that a non-conditionalizer is incoherent in that they are doing something that's bad by their own lights, even in cases with opaque evidence. In *Here's a Hand*, if your plan preferences are rationally formed, you will prefer the plan to conditionalize on EXP to all other plans for EXP , so that by failing to conditionalize on EXP in the Bad Case you are plan incoherent — you fail to implement the plan that's best by your own lights. The fact that your evidence is opaque only means that you don't know exactly what your perspective is, since your perspective is partly constituted by your evidence, so that you aren't certain of what is best by your lights.

One may also worry that this leaves the agent without guidance in such cases. However, precisely for this reason, externalists typically deny that rational norms must always give us guidance, at least in the strong sense that one's evidence must always warrant certainty about what they require.⁴⁰ Some externalists go so far as to deny that there are any norms that can always provide guidance in this sense (Williamson, 2000, ch. 4). Indeed, a similar worry arguably arises for other Bayesian arguments even without the possibility of opaque evidence, as long as one can be rationally uncertain of other factors that make up one's perspective, such as one's *credences* or *values*. For example, the *Accuracy Argument for Probabilism* (Joyce, 1998) shows that every non-probabilistic credence function is accuracy-dominated, so that for any probabilistically incoherent agent there is always a credence function that is guaranteed to be more accurate than their credences. But a non-probabilistic agent who doesn't know what their credences are can fail to know that there is a more accurate credence function, or which function that is.

⁴⁰ Hughes (2022) and Srinivasan (2015) make this point.

9. Conclusion: Bayesianism and Externalism Reconciled

We have ended up with the orthodox view that rational agents update by CONDITIONALIZATION. But we now know that this is consistent with the possibility of opaque evidence, and thus also with EXTERNALISM. The arguments for A-E CONDITIONALIZATION looked like they could circumvent the difficult question of whether one can have opaque evidence and settle the debate between externalists and their opponents by a feat of Bayesian magic, as it were. But they did so, I have argued, only by generalizing the standard arguments in the wrong way: by assuming a principle of admissibility which we have good to reject if opaque evidence is possible. However, there is an alternative way of generalizing those arguments, which forms the basis for two new arguments for CONDITIONALIZATION that apply regardless of the possibility of opaque evidence. So if life can sometimes throw opaque evidence our way, the right response is to conditionalize on it, and so remain uncertain about what we have learnt, just as we would remain uncertain about anything else that our evidence doesn't conclusively establish. This is good news for the externalist, but also for the Bayesian, who now has arguments for CONDITIONALIZATION that always apply regardless of whether opaque evidence is possible.⁴¹

Appendix: Proof of the CONDITIONAL EXPECTATION LEMMA

(Some notation will be helpful: for any acts α, β , let $\alpha_E \beta_{\neg E} = \{\langle E, \alpha \rangle, \langle \neg E, \beta \rangle\}$, i.e. the *unconditional* plan that contains the conditional plans to α if E , and to β if $\neg E$.)

Assume that $C(\cdot|E)$ is defined and that $\langle E, \phi \rangle \succeq \langle E, \psi \rangle$. REFLEXIVITY gives us $\langle \neg E, \phi \rangle \succeq \langle \neg E, \psi \rangle$. Since $\{E, \neg E\}$ is a partition, by CONDI-

⁴¹ I am particularly grateful to Michael Caie and Dmitri Gallow for invaluable advice and feedback on many drafts of this paper. For very helpful comments, I am also grateful to Branden Fitelson, Hanz MacDonald, James Shaw, two anonymous reviewers, and the editors of this journal, as well as to audiences at the 2019 University of Pittsburgh Dissertation Seminar, the 2019 meeting of the Society for Exact Philosophy, and the 2019 Formal Epistemology Workshop.

TIONAL DOMINANCE it follows that $\phi_E\phi_{-E} \succeq \psi_E\phi_{-E}$. We can then reason as follows:

By the assumption that rational preferences between unconditional plans reflect their expected value:

$$(A) \quad \sum_{\langle p, \chi \rangle \in \phi_E\phi_{-E}} \sum_{w \in p} C(w)V(\chi, w) \geq \sum_{\langle p, \chi \rangle \in \psi_E\phi_{-E}} \sum_{w \in p} C(w)V(\chi, w).$$

where $V(\chi, w)$ is the value of act χ at world w . Expanding this gives:

$$(B) \quad \sum_{w \in E} C(w)V(\phi, w) + \sum_{w \in \neg E} C(w)V(\phi, w) \geq \sum_{w \in E} C(w)V(\psi, w) + \sum_{w \in \neg E} C(w)V(\phi, w).$$

The two right summands on either side cancel out, so:

$$(C) \quad \sum_{w \in E} C(w)V(\phi, w) \geq \sum_{w \in E} C(w)V(\psi, w).$$

We can now divide both sides by $C(E)$, which is not zero since we have assumed that $C(\cdot|E)$ is defined, getting:

$$(D) \quad \sum_{w \in E} \frac{C(w)}{C(E)} V(\phi, w) \geq \sum_{w \in E} \frac{C(w)}{C(E)} V(\psi, w).$$

Then, noticing that for any $w \in E$, $C(w) = C(w \wedge E)$, and since $C(w|E) = \frac{C(w \wedge E)}{C(E)}$, we have:

$$(E) \quad \sum_{w \in E} C(w|E)V(\phi, w) \geq \sum_{w \in E} C(w|E)V(\psi, w).$$

Which, since $C(w|E) = 0$ for any $w \notin E$, gives:

$$(F) \quad \sum_{w \in \mathcal{W}} C(w|E)V(\phi, w) \geq \sum_{w \in \mathcal{W}} C(w|E)V(\psi, w).$$

Which by definition just says that:

$$(G) \quad EV_{C(\cdot|E)}\phi \geq EV_{C(\cdot|E)}\psi.$$

References

Arnold, A. (2013). Some evidence is false. *Australasian Journal of Philosophy*, 91(1), 165–172. <https://doi.org/10.1080/00048402.2011.637937>

- Barnett, D. J. (2016). Inferential justification and the transparency of belief. *Noûs*, 50(1), 184–212. <https://doi.org/10.1111/nous.12088>
- Briggs, R., & Pettigrew, R. (2020). An accuracy-dominance argument for conditionalization. *Noûs*, 54(1), 162–181. <https://doi.org/10.1111/nous.12258>
- Bronfman, A. (2014). Conditionalization and not knowing that one knows. *Erkenntnis*, 79(4), 871–892. <https://doi.org/10.1007/s10670-013-9570-0>
- Brown, P. M. (1976). Conditionalization and expected utility. *Philosophy of Science*, 43(3), 415–419. <https://doi.org/10.1086/288696>
- Caie, M. (2013). Rational probabilistic incoherence. *The Philosophical Review*, 122(4), 527–575. <https://doi.org/10.1215/00318108-2315288>
- Carr, J. R. (2017). Epistemic utility theory and the aim of belief. *Philosophy and Phenomenological Research*, 95(3), 511–534. <https://doi.org/10.1111/phpr.12436>
- Christensen, D. (1996). Dutch-book arguments de pragmatized: Epistemic consistency for partial believers. *The Journal of Philosophy*, 93(9), 450–479. <https://doi.org/10.2307/2940893>
- Comesana, J., & McGrath, M. (2016). Perceptual reasons. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 173(4), 991–1006. <https://doi.org/10.1007/s11098-015-0542-x>
- Comesaña, J., & Kantin, H. (2010). Is evidence knowledge? *Philosophy and Phenomenological Research*, 80(2), 447–454. <https://doi.org/10.1111/j.1933-1592.2010.00323.x>
- Das, N. (2019). Accuracy and ur-prior conditionalization. *The Review of Symbolic Logic*, 12(1), 62–96. <https://doi.org/10.1017/S1755020318000035>
- Easwaran, K. (2013). Expected accuracy supports conditionalization—and conglomerability and reflection. *Philosophy of Science*, 80(1), 119–142. <https://doi.org/10.1086/668879>

- Gallow, J. D. (2019). Diachronic dutch books and evidential import. *Philosophy and Phenomenological Research*, 99(1), 49–80. <https://doi.org/10.1111/phpr.12471>
- Gallow, J. D. (2021). Updating for externalists. *Noûs*, 55(3), 487–516. <https://doi.org/10.1111/nous.12307>
- Greaves, H. (2013). Epistemic decision theory. *Mind*, 122(488), 915–952. <https://doi.org/10.1093/mind/fzt090>
- Greaves, H., & Wallace, D. (2006). Justifying conditionalization: Conditionalization maximizes expected epistemic utility. *Mind*, 115(459), 607–632. <https://doi.org/10.1093/mind/fzl607>
- Hild, M. (1998). The coherence argument against conditionalization. *Synthese*, 115(2), 229–258. <https://doi.org/10.1023/A:1005082908147>
- Hughes, N. (2022). Epistemology without guidance. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 179(1), 163–196. <https://doi.org/10.1007/s11098-021-01655-8>
- Joyce, J. M. (1998). A nonpragmatic vindication of probabilism. *Philosophy of Science*, 65(4), 575–603. <https://doi.org/10.1086/392661>
- Konek, J., & Levinstein, B. A. (2019). The foundations of epistemic decision theory. *Mind*, 128(509), 69–107. <https://doi.org/10.1093/mind/fzwo44>
- Leitgeb, H., & Pettigrew, R. (2010). An objective justification of bayesianism ii: The consequences of minimizing inaccuracy. *Philosophy of Science*, 77(2), 236–272. <https://doi.org/10.1086/651318>
- Lewis, D. K. (1999). Why conditionalize? In *Papers in metaphysics and epistemology*, vol. 2 (pp. 403–407). Cambridge University Press. <https://doi.org/10.1017/CBO9780511625343.024>
- Littlejohn, C. (2011). Evidence and knowledge. *Erkenntnis*, 74(2), 241–262. <https://doi.org/10.1007/s10670-010-9247-x>
- Lord, E. (2018). *The importance of being rational*. Oxford University Press. <https://doi.org/10.1093/oso/9780198815099.001.0001>
- Maher, P. (1993). *Betting on theories*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511527326>
- Meacham, C. J. G. (2015). Understanding conditionalization. *Canadian Journal of Philosophy*, 45(5/6), 767–797. <https://doi.org/10.1080/00455091.2015.1119611>
- Paul, S. K. (2014). Diachronic incontinence is a problem in moral philosophy. *Inquiry*, 57(3), 337–355. <https://doi.org/10.1080/0020174X.2014.894273>
- Pettigrew, R. (2016). *Accuracy and the laws of credence*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198732716.001.0001>
- Rescorla, M. (2021). On the proper formulation of conditionalization. *Synthese*, 198(3), 1935–1965. <https://doi.org/10.1007/s11229-019-02179-9>
- Rizzieri, A. (2011). Evidence does not equal knowledge. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 153(2), 235–242. <https://doi.org/10.1007/s11098-009-9488-1>
- Savage, L. J. (1954). *The foundations of statistics* (Vol. 11). Wiley Publications in Statistics.
- Schoenfield, M. (2015). Bridging rationality and accuracy. *The Journal of Philosophy*, 112(12), 633–657. <https://doi.org/10.5840/jphil20151121242>
- Schoenfield, M. (2017). Conditionalization does not (in general) maximize expected accuracy. *Mind*, 126(504), 1155–1187. <https://doi.org/10.1093/mind/fzwo27>
- Srinivasan, A. (2015). Normativity without cartesian privilege. *Philosophical Issues*, 25(1), 273–299. <https://doi.org/10.1111/phils.12059>
- Teller, P. (1973). Conditionalization and observation. *Synthese*, 26(2), 218–258. <https://doi.org/10.1007/BF00873264>
- Williamson, T. (2000). *Knowledge and its limits*. Oxford University Press. <https://doi.org/10.1093/019925656X.001.0001>
- Williamson, T. (2014). Very improbable knowing. *Erkenntnis*, 79(5), 971–999. <https://doi.org/10.1007/s10670-013-9590-9>